Nonequilibrium Statistical Mechanics of the Zero-Range Process and Application to Networks

Andrew George Angel

PhD

The University of Edinburgh
2005

2

Declaration

I declare that I have composed this thesis myself; that the work is my own; and that the work has not been submitted for any other degree or professional qualification.

Signed: [redacted]

Date: 06/01/2006

Acknowledgements

This thesis would not have been possible without support and help from the following people: my first and second supervisors, Martin Evans and Mike Cates; those with whom I have worked in collaboration, David Mukamel, Erel Levine and Tom Hanney; and, of course, my parents and family.

I also thank the Carnegie Trust for the Universities of Scotland for financial support.

6

Abstract

Statistical mechanics is concerned with the study of systems with a large number of interacting constituents. Equilibrium statistical mechanics, originally introduced as a theoretical approach for thermodynamics, is well understood and a general theoretical framework exists. Nonequilibrium statistical mechanics is less well understood and no general theoretical framework currently exists. The study of nonequilibrium statistical mechanics often consists of analysing the behaviour of simple models which can exhibit interesting phenomena not expected from equilibrium systems of such simplicity.

Recently, statistical mechanics is finding application in areas that are not often strongly associated with physics, such as biology and the social sciences. These fields contain systems with many interacting constituents, but for which the concepts of equilibrium statistical mechanics are often not intuitive or applicable. Thus the study of such systems through simple nonequilibrium models is commonplace. Many such models have been proposed to represent complex networks, such as the Internet, metabolic pathway networks and social acquaintance networks. Networks represent the connections between the components of a system; the structure of these connections is often crucial to the function of the system. The recent analysis of large data sets taken from real networks has revealed high levels of organisation. The study of simple models is being used as a first step to gain insight into how and why this organisation comes about.

In this thesis a simple, stochastic, interacting particle system—the zero-range process (ZRP)—is studied with various analytical and numerical methods. In particular, the application of the ZRP and some of is generalisations to complex networks is focussed upon. The ZRP is a hopping particle model where particles hop between sites of a lattice under certain rules that depend only on the properties of the site from which the particles hop—hence the name zero-range. Through its simplicity the steady state of the ZRP can be solved, even for nonequilibrium dynamics, and yet despite its simplicity it can exhibit interesting phenomena such as condensation transitions, where a finite fraction of the total particles in the system will condense onto a single site of the lattice.

Firstly, interesting finite-size effects surrounding the condensation transition in a one-dimensional, driven version of the ZRP are studied. These take the form of discrepancies in the current-density diagram between finite and infinite systems, with

the finite behaviour resembling that seen in real traffic data.

Following this, direct applications of the ZRP to complex networks, and interesting phenomena arising from the specifics of the applications, are studied. The ZRP is applied as a model of networks and is found capable of reproducing power-law degree distributions, as observed in many real networks, at the critical point of the condensation transition. The degree is the number of connections a component of the network has. This model is then generalised to include creation and annihilation of particles or links, and this is found to exhibit critical behaviour—namely power-law particle and degree distributions—in a region of the parameter space, rather than at a critical point. The full phase diagram of this system is investigated, revealing low density and high density phases as well as subdivisions of the critical phase.

The application of the ZRP to networks is further explored through the modelling of a directed, weighted network with a multi-species ZRP. In this model there are two possible types of condensation, independent condensation of particle species onto sites and a condensation of all particle species onto a single site. These transitions are investigated along with their critical behaviour. As before the model is found to be capable of reproducing power-law distributions of various quantities, as seen in real weighted networks, at the critical points of these transitions.

Contents

Abbreviations

BA Barabási-Albert

BBV Barrat-Barthélemy-Vespignani

BRM bus-route model

ER Erdös-Rényi

LHS left hand side

MINE metabolic interaction network of Escherichia coli

NSL no self links

RG random graph

RHS right hand side

SCN scientific collaboration network

SDS single defect site

SMCN Sardinian inter-municipality commuter network

SOC self-organised criticality

SWN small world network

WAN world-wide airport network

WS Watts-Strogatz

WSF weighted scale-free

WWW world-wide web

ZRP zero-range process

Chapter 1

Introduction

The main area of physics with which this thesis is concerned is statistical mechanics. Statistical mechanics was originally introduced in the late nineteenth century as a theoretical description of thermodynamics. Statistical mechanics actually uses the fact that there is a large number of constituents in a system to its advantage. For example, it is in principle possible to model a gas by treating every atom or molecule of the gas as a particle undergoing classical Newtonian mechanics; however, this is analytically and numerically intractable for a numbers of particles much less than Avogadro's number. Statistical mechanics uses the fact that there are many constituents to make statistical predictions from simplified microscopic models of the system which typically contain the minimum amount of detail required to correctly reproduce macroscopic observable quantities.

The area of equilibrium statistical mechanics is now well developed: a general theoretical framework of statistical mechanics exists for systems that are in thermal equilibrium. Loosely speaking, this means the system is free to exchange energy with a heat bath and is at the same temperature as the bath. Under these conditions the system will attain a steady state and the correct probability distribution for the microscopic states of the model is the one which maximises the Gibbs entropy under the appropriate constraints.

Statistical mechanics and in some cases the concept of equilibrium can be extended and used to treat other systems with large numbers of constituents, often ones that are traditionally considered as being outwith the scope of physics and that do not have thermodynamic numbers of constituents. For example, statistical mechanical methods have been used to model the stock market—see [1] for a recent review. However, the concept of equilibrium is frequently not readily applicable to such systems and nonequilibrium methods are required.

Nonequilibrium statistical mechanics is less developed than its equilibrium cousin: no general theoretical framework is available for the study of nonequilibrium systems. Thus it is common to investigate very simple nonequilibrium models in order to build up a catalogue of knowledge that will hopefully lead to a general framework for the treatment of nonequilibrium systems. Many nonequilibrium systems are being driven away from equilibrium, for example by having some current of energy or mass forced through the system, but often can still attain a steady state. Such systems, known as driven diffusive systems, have proven amenable to analysis through the study of simple models in numerous cases.

A useful example of a model for driven diffusive systems is the zero-range process (ZRP) [2, 3, 4]. The ZRP is a conceptually simple Markovian [5] model in which interacting particles hop from site to site on a lattice. In many of its variants, the ZRP has several nice properties including a simple steady state which has a factorised form and the ability to display interesting phenomena such as condensation transitions. The condensation transition is characterised by a macroscopic number of particles collecting at a single site of the lattice and the factorised form for the probability makes such transitions amenable to analysis. The ZRP is studied both because it is a fecund model for fundamental nonequilibrium investigation and because it can be applied with reasonable success to model many real systems.

In this thesis, the ZRP and aspects of its behaviour are studied in detail. In particular, condensation transitions and connected phenomena are investigated, both from a fundamental point of view and from the point of view of application to real systems. In the applications, the main focus is on complex networks.

Complex networks map the structure of the interactions of interacting constituents and are composed of nodes to represent the constituents and links that connect two interacting constituents. Many systems are composed of interacting constituents and the structure of the interactions often forms a crucial backbone to the system. As such, the areas in which networks are pertinent are legion and include disciplines as diverse as biology, information technology and the social sciences. Commonly cited examples of networks that many people use everyday are transportation networks, telephone networks and the Internet.

In order to model networks in general, it is useful to think of the nodes and links themselves as interacting constituents. Thus it is appropriate to model and analyse these networks with the tools of statistical mechanics. The field of complex networks has recently enjoyed an explosion of interest in the physics community [6, 7, 8, 9] and has become an important, contemporary area of statistical physics research. Part of the reason for this surge of interest is due to generic complex behaviour seen in

measurements taken from real networks. For example, many networks have been seen to show a power-law tail in the distribution of the number of links connected to a node, suggesting some general organising principle present in networks. Many models have been proposed which try to reproduce this behaviour, and others, in the hope that they might give some insight into the way real networks are created and evolve.

The ZRP can be applied to networks in several ways; two of which are discussed in detail in this thesis. As a basic model of a network the sites of the ZRP can be thought of as nodes and the particles as ends of links; this gives a model that is capable of partially reproducing the behaviour of real networks in terms of the distribution of the number of links connected to a node. A ZRP with multiple species of particle can be used to model a weighted network in which links are assigned a value (or weight) to describe their relative importance in the function of the network. In this mapping, the particles of the ZRP represent units of weight of the links and the model is again capable of partially reproducing behaviour seen in real weighted networks; the behaviour of which is similar to, but richer than, the behaviour of unweighted networks in many respects. In both cases the ability of the ZRP to undergo condensation is instrumental in the ability of the model to reproduce realistic behaviour of the network, with the most realistic behaviour typically being seen at the critical point of the transition. In the context of networks, the condensation transition corresponds to a single node capturing a finite fraction of all the links or in the weighted case to a node or nodes capturing finite fractions of the available weight and it is at the critical point of these transitions that power-law behaviour can be observed.

In this thesis, the basic ZRP and some of its well-studied properties are introduced in Chapter 2. This includes a solution of the steady state, a general analysis of condensation transitions in the ZRP and some generalisations for which the steady state can also be found. Then in Chapter 3, an interesting finite size effect that occurs around the expected transition point and is characterised by an overshoot of the current-density diagram of the model is studied in detail. Also, connections between this and behaviour seen in data taken from real traffic models are briefly discussed. In Chapter 4, the field of complex networks is briefly reviewed from a statistical mechanics perspective. This includes discussion of several existing network models. The relation between the basic ZRP and some of these network models and the general applicability of the ZRP to networks are also discussed. In Chapter 5, a generalisation of the ZRP with creation and annihilation of particles which can show critical behaviour in a region of parameter space rather than just at a critical point is introduced and analysed. The same principle is then applied to a network model based on the ZRP; this gives a network model that is able to show realistic behaviour

in a region of the parameter space. Finally in Chapter 6 weighted networks are briefly reviewed and a model for a directed network based on a multi-species ZRP is introduced. This includes a discussion of several existing models of weighted networks and an in depth investigation of the multi-species ZRP directed, weighted network model. This model represents a novel mapping between hopping particle models and weighted networks and also displays a new kind of condensation transition that is analysed in detail. This model is seen to be capable of reproducing realistic behaviour for some weighted-specific quantities, at criticality in a similar way to the unweighted model.

Chapter 2

The Zero-Range Process

2.1 Introduction

As noted in Chapter 1, the study of nonequilibrium systems is often facilitated through simple models. Such models are currently being studied because they are both interesting in their own right and are also useful in building up a set of knowledge of nonequilibrium systems. As none yet exists, such a set of knowledge will hopefully lead to a general theoretical framework as developed and useful as that which is in place for equilibrium statistical mechanics.

Such simple models have a rich history in both equilibrium and nonequilibrium statistical mechanics. For example, the Ising model has enjoyed enormous success and publicity despite its inherent simplicity and no one would question its place in the pantheon of key developments in statistical physics.

One of the simple models that is used to study nonequilibrium systems is the zero-range process (ZRP) [2, 3, 4] and it is this model and its generalisations with which this thesis is predominantly concerned. In this chapter, the basic ZRP system and some of its previously studied generalisations are introduced, along with analysis of some of the interesting behaviour that can be seen. Further interesting behaviour, generalisations and also applications of the ZRP are discussed in the following chapters.

2.2 Basic zero-range process system

The basic ZRP is now introduced in a general, but not the most general, form and the steady state solution is given. Some of the more general forms are discussed, following the initial introduction of the model, in later sections of the chapter.

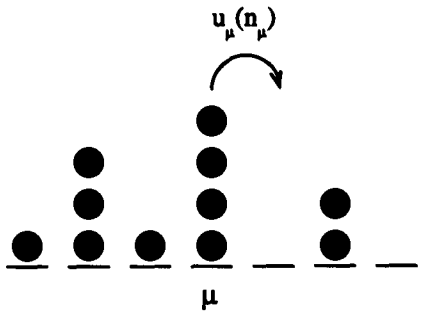


Figure 2.1: Diagram showing the layout and a possible hop move for the basic ZRP model.

2.2.1 Model definition

A one-dimensional lattice of L sites, labelled $\mu = 1 \ldots L$, upon which reside N particles is considered. Particles on site μ will hop to site $\mu+1$ with a rate $u_\mu(n_\mu)$, which depends on the site which the particles are hopping from, μ, and the number of particles on this site, n_μ. The lattice has periodic boundary conditions so a particle hopping from site L will hop to site 1. A simple diagram of a possible hop move is shown in Figure 2.1.

The hop-rates $u_\mu(n_\mu)$ correspond to the probability per unit time of a hop occurring. Thus the dynamics is conveniently simulated using a random sequential algorithm, whereby at each time step a site is chosen at random and a particle moved to the neighbouring site with probability $u_\mu(n_\mu)\Delta t$. The time interval Δt must be chosen to be small enough such that the probability of two events happening in a single time-step is negligible.

The dependence of the hop-rates on the number of particles at a site allows interaction between particles, but only those at the same site—hence the terminology zero-range. The dependence of the hop-rates on sites allows disorder in the system. As will be discussed later, these properties allow the system to exhibit condensation transitions.

2.2.2 Steady state solution

This model has a steady-state which can straightforwardly be solved [2, 3, 4]. The steady-state has the requirement that the probability current into and out of a configuration due to hops balances:

$$\sum_\mu \mathrm{P}(n_1 \ldots . n_\mu \ldots . , n_L) u_\mu(n_\mu)\theta(n_\mu) = \sum_\mu \mathrm{P}(n_1, \ldots, n_{\mu-1}+1, n_\mu - 1, \ldots, n_L) u_{\mu+1}(n_{\mu+1}+1)\theta(n_\mu) \; . \quad (2.1)$$

The left hand side of (2.1) is the total probability current out of a given configuration $\{n_\mu\} = (n_1, n_2, \ldots, n_L)$ of site occupancies of the system. The right hand side is the total probability current into this given configuration. The $\theta(n)$ are the usual Heaviside step functions [10] and are there to make clear that hops cannot cannot occur from a site with no particles.

Note that as particles can hop only to the right, a detailed balance condition—where the probability current out of a state A and into a state B is equal to the probability current from state B back into state A—cannot be satisfied. Thus the dynamics of the system are nonequilibrium in nature: a current of particles is being driven through the system. However, as shall be seen in section 2.5.1, choices for the dynamics can be made that do satisfy a detailed balance condition and have the same steady state as this system.

As an ansatz, the steady-state probability distribution is assumed to have the following factorised form:

$$\mathrm{P}(\{n_\mu\}) = \frac{1}{\mathrm{Z}(N,L)} \prod_{\mu=1}^{L} f_\mu(n_\mu) \,, \tag{2.2}$$

where the $f_\mu(n_\mu)$ are given by

$$f_\mu(n) = \begin{cases} \prod_{m=1}^{n} \dfrac{1}{u_\mu(m)} & \text{for } n > 0 \\ 1 & \text{for } n = 0 \,, \end{cases} \tag{2.3}$$

and $\mathrm{Z}(M,L)$ has been introduced to ensure that the probabilities are correctly normalised. Inserting (2.2) and (2.3) into (2.1), equating terms of the sums and cancelling common factors leads to

$$f_{\mu-1}(n_{\mu-1}) f_\mu(n_\mu) u_\mu(n_\mu) = f_{\mu-1}(n_{\mu-1}+1) f_\mu(n_\mu - 1) u_{\mu-1}(n_{\mu-1}+1) \,. \tag{2.4}$$

Rearranging a little gives

$$\frac{f_{\mu-1}(n_{\mu-1}+1)}{f_{\mu-1}(n_{\mu-1})} u_{\mu-1}(n_{\mu-1}+1) = \frac{f_\mu(n_\mu)}{f_\mu(n_\mu - 1)} u_\mu(n_\mu) \,. \tag{2.5}$$

As each side of this equation is a function of an independent variable they can be set equal to a constant and without loss of generality this constant can be taken to be equal to one. This gives the recursive relation

$$f(n_\mu) = \frac{f_\mu(n_\mu - 1)}{u(n_\mu)} \,. \tag{2.6}$$

Iterating this leads to (2.3) and thus the steady state ansatz (2.2) is valid. It is known that an irreducible Markov process with a finite state space possesses a unique steady

state, see, for example, [11, 12]. The ZRP is a Markov process; it falls under its basic definition of a stochastic process whose future evolution depends only on the present state of the system and not its past states. An irreducible Markov process is one where any state can be reached from any other given sufficient time. Thus, provided that the hop rates are chosen such that the ZRP is irreducible, an ansatz that satisfies (2.1) is the unique steady state. This argument only applies to finite systems. For a discussion of the infinite case see, for example, [13].

It should be noted that despite the factorisation property of the steady state the occupancies of the sites are correlated. This correlation comes about from the constraint that the total number of particles is conserved and comes into the equations through the normalisation quantity $\mathrm{Z}(N, L)$.

2.2.3 Normalisation

The quantity $\mathrm{Z}(N, L)$ introduced earlier to ensure the correct normalisation of the steady state probability distribution plays an analogous role to the partition function of equilibrium canonical ensemble analysis and is given by:

$$\mathrm{Z}(N, L) = \sum_{\{n_\mu\}} \prod_{\mu=1}^{L} f_\mu(n_\mu) \delta \left(\sum_\mu n_\mu - N \right) , \tag{2.7}$$

where the sums are over the occupancies of all sites and the delta function ensures that only states with the correct total number of particles contribute to the normalisation. Strictly speaking a Kronecker delta should be used, but with this the notation becomes unsightly. Thus the delta function notation is used and implicitly signifies a Kronecker delta whenever its argument is composed of integer quantities. As is usual for such a normalisation, it contains the key to much information about the system. Two quantities that can be derived from $\mathrm{Z}(N, L)$ and that are of particular interest are: the steady-state probability that a single site μ has occupation n, $\mathrm{p}_\mu(n)$; and the average hopping rate from a site in the steady state, v. The probability of finding n particles on site μ is given by:

$$\begin{aligned} \mathrm{p}_\mu(n) &= \frac{1}{\mathrm{Z}(N, L)} f_\mu(n) \sum_{\{n_\nu\} \setminus n_\mu} \prod_{\nu \neq \mu} f_\mu(n_\nu) \delta \left(\sum_{\nu \neq \mu} n_\mu - L \right) \\ &= f_\mu(n) \frac{\mathrm{Z}_\mu(N - n, L - 1)}{\mathrm{Z}(N, L)} , \end{aligned} \tag{2.8}$$

where $\mathrm{Z}_\mu(N - n, L - 1)$ is the normalisation quantity for a system with n particles and site μ removed from it and $\{n_\nu\} \setminus n_\mu$ denotes the set of occupancies of all sites except site μ. The average hopping rate out of a site, or particle current, is independent of

the site in the steady state and is given by

$$\begin{aligned} v &= \langle u_1 \rangle = \langle u_2 \rangle = \cdots \\ &= \sum_{\{n_\mu\}} u_1(n_1) \prod_{\mu=1}^{L} f_\mu(n_\mu) \delta \left(\sum_\mu n_\mu - L \right) \\ &= \frac{\mathrm{Z}(N-1, L)}{\mathrm{Z}(N, L)} , \end{aligned} \tag{2.9}$$

where $\mathrm{Z}(N-1, L)$ is the normalisation quantity for a system with $N-1$ particles and L sites. In the above, the facts $u_\mu(0) = 0$ (the hop-rate from an empty site is zero) and $u(x) f_\mu(x) = f_\mu(x-1)$ for $x \geq 1$ have been used. Thus the average hopping rate from a site is related to the logarithmic derivative of the normalisation with respect to the total particle number.

2.3 Condensation transitions

An interesting feature of the zero-range process is that it shows condensation behaviour with nonequilibrium dynamics, even on a one-dimensional lattice; something that would not be expected for a one-dimensional equilibrium system. The condensation transition is driven by the global particle density $\rho = N/L$: below a critical density, ρ_c, the system will be in an un-condensed phase, where on average the particles are evenly spread throughout the system; above the critical density, the system will be in a condensed phase where a single site holds a finite fraction of all the particles in the system. Strictly the transition is only present in the thermodynamic limit but simulations and analysis show condensation-like behaviour in large but finite systems.

Thus far, the system has been considered in a canonical ensemble, i. e. for fixed particle number. The condensation transition however is more straightforwardly analysed in a grand canonical ensemble, where the total particle number is allowed to fluctuate around some mean value. Recall that earlier both sides of (2.5) were set to the constant value of 1. Any value could have been chosen for this constant and it follows that the $f_\mu(n)$ are only defined up to some constant multiplicative factor, say z^n. Reinstating this factor, z can be interpreted as the fugacity. Thus the equivalent of the grand canonical partition function can be found in the usual way [14]: the un-normalised steady-state probability distribution is summed over all configurations with no regard as to the constraint on the total particle number; the fugacity is then chosen such that the average total particle number is equal to L. Hence the equivalent

of the grand canonical partition function, $\mathcal{Z}(z, L)$, is given by

$$\begin{aligned} \mathcal{Z}(z, L) &= \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \sum_{n_L}^{\infty} \prod_{\mu=1}^{L} z^{n_\mu} f_\mu(n_\mu) \\ &\equiv \prod_{\mu=1}^{L} F_\mu(z) \,, \end{aligned} \tag{2.10}$$

where the functions $F_\mu(z)$ are given by

$$F_\mu(z) = \sum_{n=0}^{\infty} z^n f_\mu(n) \,, \tag{2.11}$$

and play the role of grand canonical partition functions for single-site systems.

As is usual, the ensemble-average occupation for a site, $\langle n_\mu \rangle$, is given by

$$\langle n_\mu \rangle = z \frac{\partial \ln F_\mu(z)}{\partial z} \,, \tag{2.12}$$

and the constraint on the average total particle number is written:

$$\sum_{\mu=1}^{L} \langle n_\mu \rangle = N \,. \tag{2.13}$$

Combining (2.12) and (2.13) yields an equation for the global density of particles in the system, $\rho = N/L$,

$$\rho = \frac{z}{L} \sum_{\mu=1}^{L} \frac{F'_\mu(z)}{F_\mu(z)} \,, \tag{2.14}$$

where $F'(z)$ denotes the derivative of F with respect to z. The fugacity z must be chosen such that the above equation is satisfied and this equation is the key to the condensation analysis.

Now, the fugacity, z, must have a maximum value $z_{\max}$, given by the radius of convergence of (2.11). The existence of a condensation transition can be seen in the behaviour of the rhs of (2.14) as $z \nearrow z_{\max}$. As $z \nearrow z_{\max}$, the rhs of (2.14) will either converge or diverge.

If the rhs of (2.14) converges at $z = z_{\max}$, then this will give a finite critical density ρ_c, above which (2.14) cannot be satisfied for any allowed choice of z. Thus for $\rho \leq \rho_c$, the grand canonical analysis is valid and a choice of z can be made that allows the average particle number constraint to be satisfied. However, for $\rho > \rho_c$, such a choice cannot be made for z and the grand canonical analysis cannot accurately describe the system. This signifies a phase transition, which manifests itself as a condensation transition where above the critical density a single site holds a finite fraction of the total number of particles in the system.

If the rhs of (2.14) diverges at $z = z_{\max}$, then a choice of z can be made such that the equation is satisfied for any finite density by taking z arbitrarily close to $z_{\max}$. However, it is still possible to have a phase transition. Note that each term in the sum on the rhs is the mean fraction of the total density of particles on that site. If the rhs diverges through only one term in this sum diverging, then this explicitly signals a condensation transition. This is because as z is taken arbitrarily close to $z_{\max}$, the term that diverges will become arbitrarily large and the site corresponding to the diverging term must hold a finite fraction of all the particles in the system.

Thus there are two apparent mechanisms giving rise to a condensation transition: one where the rhs of (2.14) converges, which will be referred to as mechanism A; and one where the rhs of (2.14) diverges, which will be referred to as mechanism B. The choice of hop-rates will determine which mechanism takes place.

A simple, previously-studied system that exhibits mechanism A is one in which the hop-rates are the same for each site and decay with increasing particle number [15, 3, 4]. Here a condensation transition is found for hop-rates that asymptotically decay more slowly to a finite constant value β, more slowly than $\beta(1 + 2/n)$. The proof of this is given in Appendix A. Above the critical density the symmetry of the system is spontaneously broken and a randomly selected site will hold the condensate.

A simple, previously-studied system that exhibits mechanism B is one in which the hop-rates are all constant, i. e. do not depend on the number of particles, but all have different values [3, 4]. Here the site with the smallest hop-rate will hold the condensate when the critical density is exceeded. The condensation transition in this case has been likened to Bose-Einstein condensation and the conditions on the density of low energy states for condensation [14] have related conditions on the smallest hop-rates.

The critical density of the condensation transition is in principle straightforwardly found for either mechanism. For mechanism A, it is given by the value of the RHS of (2.14) at $z = z_{\max}$. For mechanism B, it is given by the largest finite value of the RHS of (2.14) as z is taken arbitrarily close to $z_{\max}$. This can often be found by the RHS of (2.14) at $z = z_{\max}$ minus the diverging term.

A system which has a mixture of decay and disorder in the hop-rates could show a condensation transition by either mechanism. Such a system will be considered in detail in Chapter 3.

For both mechanisms the condensation transition manifests itself in the following manner. Below the critical density the system is in an un-condensed phase where the particles are spread throughout the sites in a relatively even manner. In this phase the fugacity is below its maximum possible value. As the density is increased the fugacity must increase monotonically in order to satisfy (2.14). At the critical density

the fugacity attains its maximum value and the particles are spread throughout the sites in a critical manner. Above the critical density a condensate appears with a finite fraction of the total number of particles at a single site. At all other sites the particles continue to be spread out according to the critical distribution and form. Thus the condensed site takes up all excess particles as the density is increased. In the homogeneous system above the critical density, the probability distribution for the number of particles at a site takes the form of the background distribution with a peak at high n representing the possibility of a condensate at the site. This peak has area $1/L$ so that the probability of having a condensate at one of the sites is 1.

The grand canonical analysis allows the easy calculation of the probability distribution for the number of particles at a site. As in the grand canonical ensemble there are no constraints on the total number of particles, the steady state factorises into independent factors—there is no delta function present to couple the factors as for the canonical ensemble. Thus in the grand canonical ensemble the distribution for the number of particles on site μ is given by

$$\mathrm{p}_\mu(n) = \frac{f_\mu(n) z^n}{F_\mu(z)} \,. \tag{2.15}$$

This equation, of course, is only valid when the grand canonical analysis is valid, i. e. when the system is not in the condensed phase. However, it also proves useful in the condensed phase, as inserting the maximum value of the fugacity into the above equation gives the critical, background distribution that correctly describes the system except for the condensate.

In the homogeneous case, the critical distribution of particles can take on a power-law tail. This is realised when the hop rates behave asymptotically as $u(n) \sim \beta(1 + b/n)$, with $b > 2$. Inserting this into (2.3) yields $f(n) \sim \beta^{-n} n^{-b}$. Then inserting this and the maximum value of the fugacity (which is β in this case) into (2.15) gives $\mathrm{p}(n) \sim n^{-b}$ and the probability of particles at a site has a power-law tail. This property of the system proves important when applications to networks are discussed in later chapters.

In this chapter. only the statics of the condensation transition have been discussed. Further discussion of condensation transitions with an emphasis on the dynamics of the transitions can be found in [16, 17].

2.4 Connection between the canonical and grand-canonical ensembles

To motivate the validity of the grand-canonical analysis presented above, the connection between the two ensembles is now briefly discussed. It is possible to get to the grand-canonical density equation (2.14), by starting from the canonical normalisation $\mathrm{Z}(N, L)$ (2.7) and inserting the integral representation of the delta function

$$\delta(x-a) = \oint \frac{\mathrm{d}z}{2\pi i} \frac{z^x}{z^{a+1}} \,, \tag{2.16}$$

which yields

$$\begin{aligned} \mathrm{Z}(N, L) &= \oint \frac{\mathrm{d}z}{2\pi i} z^{-(N+1)} \prod_{\mu} \sum_{n_\mu} z^{n_\mu} f_\mu(n_\mu) \\ &= \oint \frac{\mathrm{d}z}{2\pi i} z^{-(N+1)} \prod_{\mu} F_\mu(z) \end{aligned} \tag{2.17}$$

where the functions $F_\mu(z)$ have the same form as those introduced in the grand canonical ensemble (2.11). Thus the canonical normalisation may be written

$$\mathrm{Z}(N, L) = \oint \frac{\mathrm{d}z}{2\pi i} z^{-(N+1)} \mathcal{Z}(z, L) \,. \tag{2.18}$$

Note that the variable z introduced is just a dummy integration variable at this point. Now, (2.17) can be written

$$\mathrm{Z}(N, L) = \oint \frac{\mathrm{d}z}{2\pi i} \frac{1}{z} \exp\left[-N \ln z + \sum_{\mu=1}^{L} \ln F_\mu(z) \right] \,. \tag{2.19}$$

For large N and L this integral will be dominated by the maximum of the exponential and the integral can be evaluated by using the saddle point method [18]. The saddle point is found by solving

$$\frac{\partial}{\partial z} \left[-N \ln z + \sum_{\mu=1}^{L} \ln F_\mu(z) \right] = 0 \,, \tag{2.20}$$

which gives the saddle point equation

$$\rho = \frac{z}{L} \sum_{\mu=1}^{L} \frac{F'_\mu(z)}{F_\mu(z)} \,. \tag{2.21}$$

Thus the density equation of the grand canonical analysis (2.14) is recovered and the grand-canonical and canonical results are equivalent in the thermodynamic limit.

Similar methods can be used to show the equivalence of the particle current v, as found in the canonical analysis (2.9), and the fugacity z, as introduced in the

grand-canonical analysis, in the thermodynamic limit. Writing (2.19) in the general form

$$\oint \mathrm{d}z\, g(z)\, \exp\left[Nh(z)\right] , \tag{2.22}$$

where $\mathrm{g}(z) = 1/z$ and $\mathrm{h}(z) = -\ln(z) + \sum_\mu \ln F_\mu(z)$, and noting that this has the leading asymptotic behaviour for $N \gg 1$ (assuming an appropriate saddle point can be found [18])

$$g(z_0) \exp\left[Nh(z_0)\right] e^{i\phi} \left(\frac{2\pi}{N|h''(z_0)|} \right)^{1/2} , \tag{2.23}$$

where z_0 is the location of the saddle point and ϕ is the angle of steepest descent through the saddle point, yields

$$v = \frac{\mathrm{Z}(N-1, L)}{\mathrm{Z}(N, L)} \sim z_0 \, . \tag{2.24}$$

Now, z_0 is the position of the saddle point and from (2.21) this is the same as the fugacity from the grand-canonical analysis. Thus the fugacity is equivalent to the average hop rate from a site, or particle current, in the thermodynamic limit.

With this relation (2.24) the behaviour of the current can easily be seen as it is the same as the behaviour of the fugacity. Below the critical density the current is monotonically increasing and at and above the critical density the current attains a finite constant value.

2.5 Generalisations

The zero-range process defined above in Section 2.2.1 is already quite general, however, there are a few more important generalisations that also need to to be considered.

2.5.1 General lattices

Previously the model was defined on a one-dimensional ring lattice with hops to the rightmost adjacent site only, i. e. a particle hops out of a site μ with rate $u_\mu(n_\mu)$ and moves to site $\mu + 1$ with probability 1. The zero-range process can also be solved on more general lattices, i. e. one on which a particle hops out of a site μ with rate $u_\mu(n_\mu)$ and then jumps to a site ν with probability $\mathrm{W}_{\nu\mu}$ [3, 4]. For the steady state to be soluble and have a factorised form it is required that a single particle performing a random walk on the lattice under the set of transition probabilities $\{\mathrm{W}_{\nu\mu}\}$ has a unique steady-state solution. For this to be true, it is necessary that

$$\sum_\nu \mathrm{W}_{\nu\mu} = 1 \, , \tag{2.25}$$

for all μ and the steady-state probability distribution of the single particle random walker being found at site μ is then given through

$$\mathrm{s}_\mu = \sum_{\nu \neq \mu} \mathrm{W}_{\mu\nu}\, \mathrm{s}_\nu \ . \tag{2.26}$$

Some simple cases where a unique steady state solution for this exists have been discussed in [2, 3]. Again the factorisation ansatz for the steady-state probability distribution is made

$$\mathrm{P}(\{n_\mu\}) = \frac{1}{\mathrm{Z}(N,L)} \prod_{\mu=1}^{L} f_\mu(n_\mu) \ . \tag{2.27}$$

Where the f_μ now have a slightly different form incorporating the steady-state solution of the random walker problem (2.26):

$$f_\mu(n) = \begin{cases} \displaystyle\prod_{m=1}^{n} \frac{\mathrm{s}_\mu}{u_\mu(m)} & \text{for } n > 0 \\ 1 & \text{for } n = 0 \ . \end{cases} \tag{2.28}$$

This can be proven in a similar way as before. Starting with the steady-state requirement that the probability current out of a state is equal to the probability current into that state,

$$\begin{aligned} &\sum_\mu \mathrm{P}(n_1, \ldots, n_\mu, \ldots, n_\nu, \ldots, n_L) u_\mu(n_\mu)\theta(n_\mu) = \\ &\sum_\mu \sum_{\nu \neq \mu} \mathrm{P}(n_1, \ldots, n_\mu - 1, \ldots, n_\nu + 1, \ldots, n_L) \mathrm{W}_{\mu\nu} u_\nu(n_\nu + 1)\theta(n_\mu) \ , \end{aligned} \tag{2.29}$$

inserting (2.27) into (2.29), equating terms in the sums and cancelling common factors yields

$$f_\mu(n_\mu)u_\mu(n_\mu) = \sum_{\nu \neq \mu} f_\mu(n_\mu - 1) \frac{f_\nu(n_\nu + 1)}{f_\nu(n_{nu})} \mathrm{W}_{\mu\nu} u_\nu(n_\nu + 1) \ . \tag{2.30}$$

Inserting (2.28) into (2.30) recovers the steady-state solution of the random walker problem (2.26), thus the steady-state solution is proven for the arbitrary lattice.

Some particularly simple cases of this are: one-dimensional lattices where particles move to either neighbouring site with equal probability when hopping; and fully connected lattices where particles move to any lattice site other than the one they hop from with equal probability. In both of these cases the random walker is equally likely to be found at any site in the steady state, i. e. s_μ is a site independent constant. Thus these cases share exactly the same steady-state probability distribution as for the model with hops only to the right on a ring (2.2). It is also interesting to note that both these systems satisfy a detailed balance condition and hence can be considered to be equilibrium systems; the ZRP can give the same steady state for systems both in and out of equilibrium. This point has been discussed in [3].

2.5.2 Multiple species of particles

Systems with more than one species of particle are becoming increasingly popular for several reasons: their study can add to the general catalogue of knowledge of nonequilibrium systems; the addition of extra species has been known to create new and interesting physics in some systems; and also because they can expand the application possibilities for models. Recently, generalised zero-range processes with two or more species of particle have been investigated [19, 20, 21, 22, 23, 4]. The most general case is where each species of particle has a dynamics that depends on the number of each and every species of particle at a site and the dynamics of each species can be different, i. e. the hop rates depend on the set of occupancies of each species and can be different for each species.

To begin with the case of only two species of particle is discussed. For simplicity the particles are restricted to hopping to the rightmost nearest neighbour on a one-dimensional lattice with periodic boundary conditions. Generalisations to other lattices can be made in a similar fashion as with the single species zero-range process in Section 2.5.1.

Consider a ring lattice of L sites labelled $\mu = 1 \ldots L$, upon which reside N particles of species A and M particles of species B, say. Particles of species A on site μ will hop from this site with rate $u_\mu(n_\mu, m_\mu)$, which depends on the site, μ, the number of particles of species A on this site, n_μ, and the number of particles of species B on this site, m_μ. Particles of species B on site μ will hop from this site with rate $v_\mu(n_\mu, m_\mu)$, which again depends on the site and the numbers of particles of species A and B on that site.

Under certain constraints on these hop-rates [19, 22], which are given by equation (2.41), the factorisation property of the single-species zero-range process is retained for this two-species generalisation. The steady-state probability of a configuration $(\{n_\mu\}\,;\{m_\mu\})$, is given by

$$\mathrm{P}\left(\{n_\mu\}\,;\{m_\mu\}\right) = \frac{1}{\mathrm{Z}(N,M,L)} \prod_{\mu=1}^{L} f_\mu(n_\mu, m_\mu)\ . \tag{2.31}$$

In this multi-species case, there are several forms that the f_μ functions can take, one of which is

$$f_\mu(n,m) = \begin{cases} \displaystyle\prod_{i=1}^{n} \frac{1}{u_\mu(i,m)} \prod_{j=1}^{m} \frac{1}{v_\mu(0,j)} & \text{for } n,m > 0 \\ 1 & \text{for } n = m = 0\ . \end{cases} \tag{2.32}$$

The requirements on the hop rates for a factorised steady state are basically such that all possible forms for (2.32) are equivalent. Again a normalisation quantity

$\mathrm{Z}(N, M, L)$ has been introduced, which is analogous to the canonical partition function of equilibrium statistical mechanics and is given by

$$\mathrm{Z}(N, M, L) = \sum_{\{n_\mu\},\{m_\mu\}} \prod_{\mu=1}^{L} f_\mu(n_\mu, m_\mu) \delta\left(\sum_{\mu=1}^{L} n_\mu - N\right) \delta\left(\sum_{\mu=1}^{L} m_\mu - M\right) , \quad (2.33)$$

where there are now two delta functions to separately constrain the number of particles of both species.

The condition on the hop-rates for factorisation can be found and the steady state proven in much the same way as for the single-species zero-range process. The starting point is, as always, the condition that the probability current out of a configuration due to hops should be equal to the probability current into this configuration due to hops:

$$\begin{aligned}\sum_\mu \mathrm{P}\begin{pmatrix} n_1, \ldots, n_{\mu-1}, n_\mu, \ldots, n_L \\ m_1, \ldots, m_{\mu-1}, m_\mu, \ldots, m_L \end{pmatrix} [u_\mu(n_\mu, m_\mu)\theta(n_\mu) + v_\mu(n_\mu, m_\mu)\theta(m_\mu)] = \\ \sum_\mu \Bigg[\mathrm{P}\begin{pmatrix} n_1, \ldots, n_{\mu-1}+1, n_\mu - 1, \ldots, n_L \\ m_1, \ldots, m_{\mu-1}, m_\mu, \ldots, m_L \end{pmatrix} u_{\mu-1}(n_{\mu-1}+1, m_{\mu-1})\theta(n_\mu) \\ + \mathrm{P}\begin{pmatrix} n_1, \ldots, n_{\mu-1}, n_\mu, \ldots, n_L \\ m_1, \ldots, m_{\mu-1}+1, m_\mu - 1, \ldots, m_L \end{pmatrix} v_{\mu-1}(n_{\mu-1}, m_{\mu-1}+1)\theta(m_\mu) \Bigg] . \quad (2.34)\end{aligned}$$

As before, terms of the sums are equated, the factorised form (2.31) inserted and common terms cancelled, but this time the further step of equating terms that contain species A hop-rate $u(n, m)$ and species B hop-rate $v(n, m)$ is taken. Doing so and re-arranging a little gives

$$\frac{f_\mu(n_\mu, m_\mu)}{f_\mu(n_\mu - 1, m_\mu)} u_\mu(n_\mu, m_\mu) = \frac{f_{\mu-1}(n_{\mu-1}+1, m_{\mu-1})}{f_{\mu-1}(n_{\mu-1}, m_{\mu-1})} u_{\mu-1}(n_{\mu-1}+1, m_\mu) , \quad (2.35)$$

and

$$\frac{f_\mu(n_\mu, m_\mu)}{f_\mu(n_\mu, m_\mu - 1)} v_\mu(n_\mu, m_\mu) = \frac{f_{\mu-1}(n_{\mu-1}, m_{\mu-1}+1)}{f_{\mu-1}(n_{\mu-1}, m_{\mu-1}} v_{\mu-1}(n_{\mu-1}, m_{\mu-1}+1) . \quad (2.36)$$

In the same way as before each side of these equations (2.35), (2.36) can be equated to a constant as they depend on independent variables and this constant can be set to be 1 without loss of generality. The following two recursion relations are then found:

$$f_\mu(n, m) = \frac{f_\mu(n-1, m)}{u_\mu(n, m)} \quad (2.37)$$

$$f_\mu(n, m) = \frac{f_\mu(n, m-1)}{v_\mu(n, m)} . \quad (2.38)$$

From this it is seen that there is a difference from the single-species case in that, in general, the order of iteration matters: Two different relations can be found between

$f_\mu(n,m)$ and $f_\mu(n-1,m-1)$ depending on whether (2.37) is applied before or after (2.38). The two possible relations are:

$$f_\mu(n,m) = \frac{f_\mu(n-1,m)}{u_\mu(n,m)} = \frac{f_\mu(n-1,m-1)}{u_\mu(n,m)v_\mu(n-1,m)} , \tag{2.39}$$

on applying (2.37) then (2.38); and

$$f_\mu(n,m) = \frac{f_\mu(n,m-1)}{v_\mu(n,m)} = \frac{f_\mu(n-1,m-1)}{u_\mu(n,m-1)v_\mu(n,m)} , \tag{2.40}$$

on applying (2.38) then (2.37). Thus, for the factorised steady-state given by (2.31) and (2.32) to be valid, these two relations must be equivalent, placing the following constraint on the hop-rates $u(n,m)$, $v(n,m)$

$$\frac{u_\mu(n,m)}{u_\mu(n,m-1)} = \frac{v_\mu(n,m)}{v_\mu(n-1,m)} . \tag{2.41}$$

With this constraint (2.41), the order in which the iteration relations (2.37) and (2.38) are applied does not matter and iterating relation (2.37) n times followed by iterating (2.38) m times recovers the form of the $f_\mu(n,m)$ given in (2.32). Thus the steady-state for the two-species system is proven.

It should be noted that the constraint on the hop rates (2.41) makes the system not as general as it may first have appeared, if the factorisation property of the steady state is to hold: the hop rates of the species must depend on one another in a prescribed way. However, the system is still capable of showing interesting behaviour. In particular, the two-species system opens up new condensation possibilities and condensation transitions have been seen where a 'weak' condensate of one species induces a 'strong' condensate of the other [19, 20].

Generalisations to more than two-species is also possible [22]. Here there will be a set of constraints similar in form to (2.41); one for each possible pairing of species [22]

$$\frac{u^i(\mathbf{n})}{u^i(\mathbf{n}_{\setminus j}, n_j - 1)} = \frac{u^j(\mathbf{n})}{u^j(\mathbf{n}_{\setminus i}, n_i - 1)} , \tag{2.42}$$

where u^i is the hop rate for species i, $\mathbf{n}$ is the set of the occupations of all species at a site and $\mathbf{n}_{\setminus i}$ is the set of occupations of all species except species i at a site. A specific example of a multi-species system will be discussed in detail in Chapter 6.

2.6 Numerical study of the zero-range process

The zero-range process and many of its generalisations are amenable to study with numerical methods. In particular, dynamical Monte Carlo simulations and numerical

evaluation of the partition function are useful to study condensation and related behaviour in the ZRP. Of course numerical studies can only be done on finite systems and as such true condensation behaviour can never be observed in this way. However, for large-but-finite systems condensation-like behaviour can often be clearly distinguished and it is useful to study this and compare to the expected thermodynamic behaviour. If an appropriate order parameter is defined, e. g. the total number of particles on the most occupied site divided by the total number of particles in the system, then often it will be measured as zero far below the critical density and significantly non-zero far above the critical density. Thus, although the behaviour close to the transition point may be blurred, distinct putative non-condensed and condensed phases can be observed in finite systems.

2.6.1 Monte Carlo simulation

Monte Carlo simulations of the zero-range process have already been briefly discussed when motivating the basic dynamics of a ZRP. The simulation of the ZRP is in general very simple to do and is discussed in a little more detail in this section.

The usual way to simulate the ZRP is to use a simple random sequential algorithm, whereby at each timestep a site is selected at random and a particle hopped from this site to another with an appropriate probability. Parallel updating schemes do exist and these tend to be much faster; however, these give different results to the random sequential algorithm and require a different theory [24].

One of the reasons for the ease of simulation is the fact that the configuration of a ZRP can quite easily be stored in an array of integer variables. For the single-species ZRP a one-dimensional integer array of L elements, with each element representing a site and the entry of each element being the number of particles on the corresponding site, holds a complete description of the system. For multi-species systems a two-dimensional array is used, with the second dimension used to distinguish between the species of particle.

With this storage method and for random sequential updating the simulation is divided into time steps of length Δt. At each step the following update procedure takes place

- A site μ is selected at random by generating a random number in the range 0 to $L-1$ and selecting the relevant element of the array.

- A particle is removed from this site with probability $u_\mu(n_\mu)\Delta t$, i. e. the value of element $\mu-1$ of the configuration array is decremented by 1 if a random number

in the range 0 to 1 is less than $u_\mu(n_\mu)\Delta t$. Note that Δt is chosen such that $u\Delta t$ is less than or equal to 1.

- If a particle was removed then one is added to site ν with probability $\mathrm{W}_{\nu\mu}$, i. e. the value of element $\nu - 1$ of the configuration array is incremented by 1 if a random number generated in the range 0 to $\sum_x \mathrm{W}_{x\mu}$ falls between $\sum_{x=1}^{\nu-1} \mathrm{W}_{x\mu}$ and $\sum_{x=1}^{\nu} \mathrm{W}_{x\mu}$.

This process is then repeated until the end of the simulation. To ensure that any measurements are taken in the steady state, it is usual to allow the simulation to run until it is deemed to have settled and then measure and average various quantities of interest over the remaining steps. This is most commonly achieved by measuring the value of a property at evenly spaced intervals, storing it (often cumulatively) and taking the mean at the end. This gives a simple time average which should be equivalent to the ensemble average.

An example of the code to simulate a ZRP where the the average fraction of particles on the most occupied site and the average hopping rate from a site are calculated is given in Appendix C.

2.6.2 Numerical computation of the partition function

A useful feature of the ZRP and its generalisations is that the normalisation $\mathrm{Z}(N, L)$ (2.7), the equivalent of the canonical partition function, can often be calculated numerically and this computation can be exact up to the precision of the machine being used. The numerically evaluated $\mathrm{Z}(N, L)$ can then be used to get exact forms for quantities like the probability distribution for a single site's occupancy, $\mathrm{p}_\mu(n)$ (2.8), and the particle current, v (2.9), as these are both directly related to the normalisation. Due to the specifics of the calculation it can only be used for finite systems and while this means it cannot be used to fully check thermodynamic predictions, it is ideal for studying finite-size effects and comparing with simulations of finite systems.

The normalisation is evaluated using a recursion relation; the derived form of the single site probability distribution (2.8) being the key to finding this relation. Given (2.8) and the fact that the sum over this single site probability distribution must be equal to one, it is found that

$$\mathrm{Z}(N, L) = \sum_{n=0}^{N} f_\mu(n) \mathrm{Z}_\mu(N - n, L - 1) \,, \tag{2.43}$$

where $\mathrm{Z}_\mu(N - n. L - 1)$ is the normalisation of a system with site μ and n particles removed. Now

$$\mathrm{Z}_{\{\mu\}\backslash\nu}(x, 1) = f_\nu(x) \,, \tag{2.44}$$

where $\mathrm{Z}_{\{\mu\}\backslash\nu}(x,1)$ is the normalisation for a system with all sites except site ν having been removed and with x particles in it. Thus the normalisation for a system with N particles and L sites can be successively built up from the functions $f_\mu(n)$ (2.3) and the normalisations of systems with fewer sites and particles up to the number required for the final normalisation, starting from the normalisations for a single-site system.

In order to avoid computational divergences the actual algorithm used when computing the normalisation will be something like

$$\ln \mathrm{Z}(N,L) = \ln \mathrm{Z}_\mu(N,L-1) + \ln\left(\sum_{n=0}^{N} \exp\left[\ln f_\mu(n) + \ln \mathrm{Z}_\mu(N-n,L-1) - \ln \mathrm{Z}_\mu(N,L-1)\right]\right) . \quad (2.45)$$

By taking logarithms in this way, overflow errors are avoided. Although exponentiation operations are present, they only have differences of logarithms as their argument and so do not contribute to overflow errors. If large enough system sizes are used then overflow errors are unavoidable using standard double precision variables. Algorithms for doing mathematical operations to any precision do exist [25] , but these tend to increase run-time to impractical levels. Overflow also tends to happen when the density of a system passes a certain level. For large enough system sizes this density can be quite low, although often run-times will approach impractical levels before this becomes a serious problem.

An example of the code to calculate the normalisation for the homogeneous ZRP is given in Appendix D.

2.7 Summary

In this chapter, a simple model where particles hop from site to site on a lattice—the zero-range process—and its analysis have been introduced. This included a discussion and analysis of condensation transitions, where particles of the system tend to collate at the same site of the lattice. Generalisations of the model to more general lattices and multiple species of particle were also discussed. Finally, descriptions of how to simulate the zero-range process and how to numerically calculate the equivalent of a partition function for the system exactly were given.

Chapter 3

Single Defect Site: The Overshoot

3.1 Introduction

The ZRP introduced in Chapter 2 is a simple model that has been used, albeit sometimes under different names, to model various non-equilibrium systems, for example polymer dynamics [26], sandpile dynamics [27] and glasses [28]. Despite its simplicity the ZRP can exhibit non-trivial behaviour, such as condensation transitions which have been discussed in detail in Section 2.3. In this chapter an interesting finite size effect that happens around the condensation transition is studied and its relation to behaviour seen in vehicular traffic is briefly discussed.

3.2 The 'overshoot'

In this section, the interesting finite-size effect—known as the 'overshoot'—that occurs in the ZRP and some other systems is introduced. A particular case of the ZRP system defined in Section 2.2.1, i. e. a one-dimensional ring of L sites, upon which reside N particles which hop from site μ to site $\mu+1$ with rate $u_\mu(n_\mu)$, where n_μ is the number of particles on site μ, will be used to observe and discuss this effect.

Specifically, a case of the system where the hop rates are chosen such that the system will undergo a condensation transition at finite density in the thermodynamic limit, as discussed in Section 2.3 is considered. As discussed in Chapter 2 Section 2.4, an interesting characteristic of such a system is the behaviour of the particle current v (defined as the average hop rate from a site in the steady-state and given by equation (2.9)) with varying global particle density ρ. For the condensation transitions

discussed in Section 2.3 the current has two distinct regions of behaviour: below the critical density, ρ_c, the current is monotonically increasing with density; above ρ_c, the current has a constant value. This behaviour is shown as the solid line in Figure 3.1.

In large but finite systems, the current density diagram can show a significant deviation from this behaviour. Below the critical density the current is still monotonically increasing with density and often shows excellent agreement with the expected thermodynamic current. However, above the critical density the current often goes above its expected thermodynamic saturation value, before tending to this value from above at higher density. The finite system current 'overshoots' the expected thermodynamic saturation value, as shown in Figure 3.1.

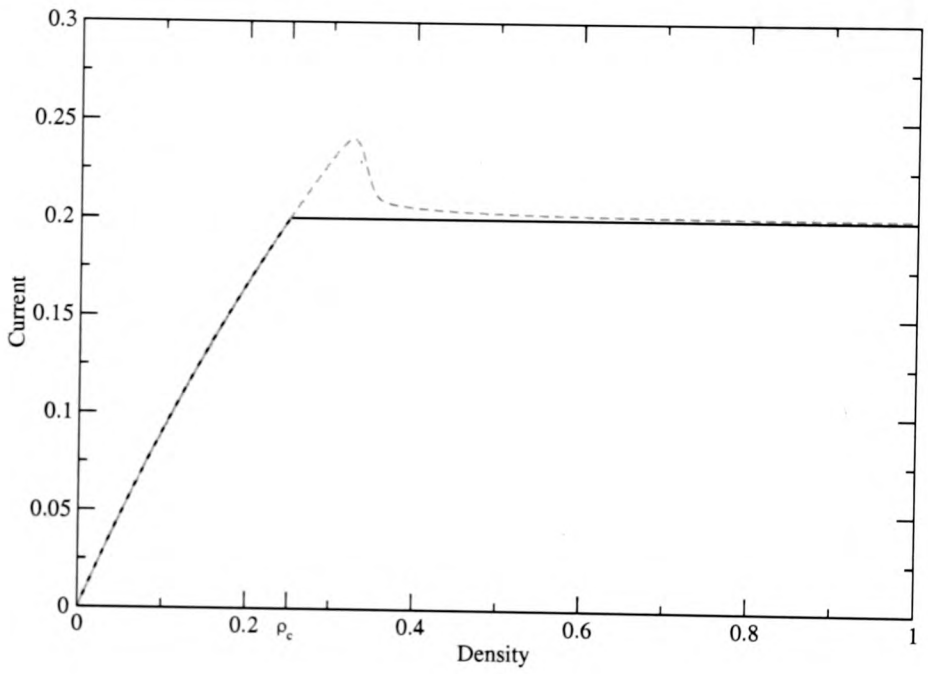


Figure 3.1: An example of the overshoot in the particle current versus global particle density diagram. The solid black line shows the expected thermodynamic behaviour. The dashed grey line shows the behaviour of a large-but-finite system taken from an exact numerical evaluation of the current. In the thermodynamic limit, the current is equivalent to the fugacity, z. The data shown were taken from the SDS system of Section 3.3 with the slow site hop rate $u_1(n) = 0.2(1 + 4/n)$. The critical density for the system in the thermodynamic limit, ρ_c, is equal to 0.25 and is marked on the graph; this coincides with the point at which the current first deviates from the expected thermodynamic behaviour.

This effect is quite general: it is seen for many systems in which the hop rates have forms that decay asymptotically to some constant finite value. For example, the overshoot can be seen in: homogeneous systems with hop rates of the form $u(n) = 1 + b/n^\sigma$, with $b > 2$ for $\sigma = 1$ and $b > 0$ for $0 < \sigma < 1$ (where condensation takes

place via mechanism A) ; and inhomogeneous systems with hop rates of the form $u_\mu(n) = q_\mu(1 + b/n^\sigma)$, with $b, \sigma > 0$ and $\{q_\mu\}$ a set of constants with one sufficiently smaller than the rest (where condensation can take place via both mechanism A and mechanism B) Figure 3.2 shows some examples of overshoots from these systems, the data have been scaled so that the expected saturation value is equal to 1. Note that the homogeneous system with $\sigma > 1$ (dot-dashed line) is tending to the saturation value very slowly.

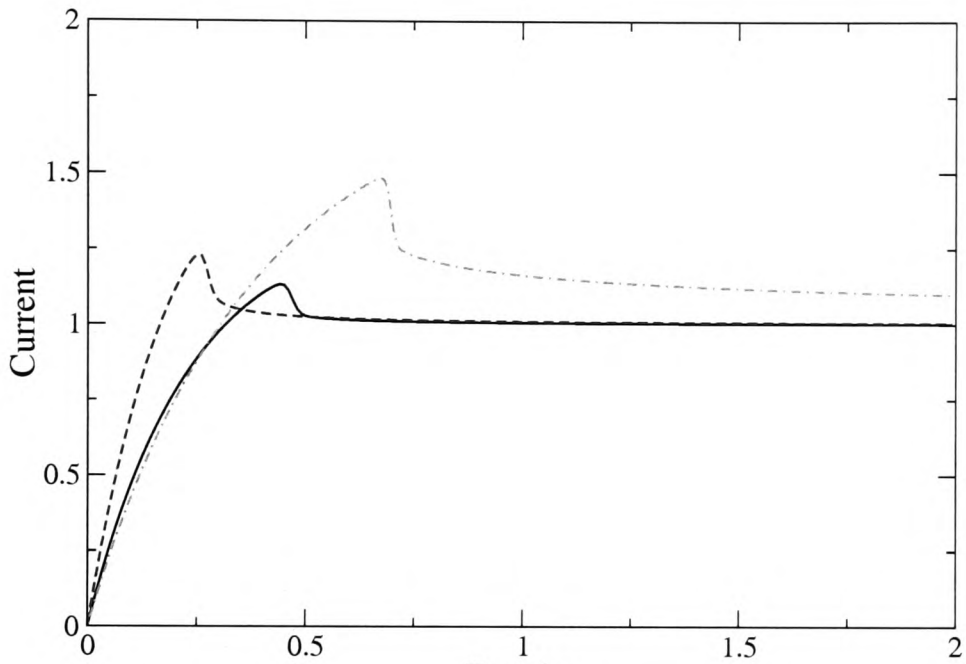


Figure 3.2: Examples of overshoots in the current-density diagram for several different ZRP systems. Solid line is from an inhomogeneous system with hop rates $u_1(n) = 0.85(1+4/n)$ and $u_{\mu\neq 1} = 1+4/n$. Dashed line is from a homogeneous system with hop rate $u(n) = (1+8/n)$. Dot-dashed line is from a system with hop rate $u(n) = 1+4/n^{0.5}$. All data have been scaled such that the expected infinite saturation value is equal to 1. It is not always possible to evaluate the critical density, but in all cases where it can be evaluated it closely matches the point at which the finite system current first crosses the expected saturation value. The dot-dashed line represents data from a system where it is difficult to calculate the critical density. For the other two lines the critical densities coincide very closely with the point at which the current first crosses the saturation value.

For some of these example systems, although they appear simple, the overshoot is actually somewhat difficult to analyse. A particularly simple model which exhibits the desired overshoot behaviour, can undergo condensation via mechanisms A and B and is amenable to analysis is analysed in detail in the following sections.

3.3 Single-defect-site system

In order to analyse the overshoot effect in detail, a very simple system will be analysed. The general basic model defined in Section 2.2.1 will be used with a particular choice of hop rates: All rates except one are constant and taken as 1, while one rate only, taken to be from site $\mu = 1$ for simplicity, has the form $p(1 + b/n)$ with $0 < p < 1$ and $b > 0$, i. e. the system is defined through the following hop rates

$$u_1(n) = p(1 + b/n) \tag{3.1a}$$

$$u_{\mu \neq 1} = 1 \, , \tag{3.1b}$$

both for $n > 0$. Thus the system has a single 'defect' site and is referred to as the single-defect-site (SDS) system.

3.3.1 Grand-canonical analysis

As with the standard ZRP, analysis of the condensation transition is most straightforwardly done using a grand-canonical formalism, where the total number of particles in the system is only constrained to a mean value and allowed to fluctuate. The grand canonical condensation analysis of section 2.3 is now followed in detail for the SDS system with the choice of hop rates as given above in (3.1). With these hop rates, the functions $f_\mu(n)$ introduced in (2.2), (2.3) are found to be, for $n > 0$

$$f_1(n) = \frac{1}{p^n} \frac{n!}{(1+b)_n} \tag{3.2a}$$

$$f_\mu(n) = 1 \qquad \text{for } \mu > 1 \, , \tag{3.2b}$$

where $(a)_n$ is the Pochhammer symbol defined by

$$\begin{aligned} (a)_m &\equiv a(a+1)(a+2)\cdots(a+m-1) \\ (a)_0 &\equiv 1 \, . \end{aligned} \tag{3.3}$$

The $F_\mu(z)$ generating functions (2.11) take simple exact forms in the SDS system

$$F_1(z) = \sum_{n=0}^{\infty} \left(\frac{z}{p}\right)^n \frac{n!}{(1+b)_n} = {}_2\mathrm{F}_1(1, 1; 1+b; z/p) \, , \tag{3.4}$$

where ${}_2\mathrm{F}_1(a, b; c; x)$ is the hypergeometric function [29] defined by

$${}_2\mathrm{F}_1(a, b; c; x) = \sum_{m=0}^{\infty} \frac{(a)_m (b)_m}{(c)_m} \frac{x^m}{m!} \, . \tag{3.5}$$

To see the equivalence between (3.4) and (3.5) note that $(1)_m = m!$. The $F_\mu(z)$ for $\mu > 1$ are simple geometric sums

$$F_\mu(z) = F(z) = \sum_{m=0}^{\infty} z^n = \frac{1}{1-z} \,. \tag{3.6}$$

The density constraint equation (given by 2.14) takes the following form for this system:

$$\rho = \frac{z}{L}\left(\frac{F_1'(z)}{F_1(z)} + (L-1)\frac{F'(z)}{F(z)}\right) \,. \tag{3.7}$$

Noting that [29]

$$F_1'(z) = \frac{\mathrm{d}}{\mathrm{d}z}\,{}_2\mathrm{F}_1(1,1;1+b;z/p) = \frac{1}{p(1+b)}\,{}_2\mathrm{F}_1(2,2;2+b;z/p) \,, \tag{3.8}$$

and inserting $F_1(z)$, $F_1'(z)$, $F(z)$ and $F'(z)$ into (3.7) gives

$$\begin{aligned} \rho &= \frac{z}{L}\frac{{}_2\mathrm{F}_1(2,2;2+b;z/p)}{p(1+b)\,{}_2\mathrm{F}_1(1,1;1+b;z/p)} + \left(1-\frac{1}{L}\right)\frac{z}{1-z} \\ &= \frac{\langle n_1(z)\rangle}{L} + \left(1-\frac{1}{L}\right)\langle n(z)\rangle \,. \end{aligned} \tag{3.9}$$

Here $\langle n_1(z)\rangle$ is the ensemble average of the number of particles on the defect site and $\langle n(z)\rangle$ is the ensemble average of the number of particles on any of the non-defect sites; the average occupancies of all the non-defect sites are identical. Equation (3.9) is a key part of the analysis: as in the general analysis of Section 2.3, it is the convergence properties of this equation that determine the existence and type of condensation transitions.

Despite being extremely simple, the SDS system is capable of undergoing condensation via either of the two mechanisms, A and B, identified in Section 2.3. As this specific choice of system allows for a more detailed analysis than the general case, the two mechanism will be reconsidered here at length. Recall that the maximum value of z is given by the smallest radius of convergence from the set of the $F_\mu(z)$ functions. In the case of the SDS system this value is p. The convergence properties of the RHS of (3.9) at this maximum value of z will determine by which mechanism condensation takes place.

The second term of the RHS of (3.9) will always converge at the maximum value of z, thus the convergence of the first term determines which mechanism takes place. The convergence properties of the hypergeometric function are known [29], ${}_2\mathrm{F}_1(a,b;c;x)$ converges as $x \nearrow 1$ if $c > a+b$. Thus the RHS of (3.9) will converge only if $b > 2$ for the SDS system; the system shows condensation via mechanism A within this range and condensation via mechanism B outwith it.

For $b > 2$ condensation takes place via mechanism A. The RHS of (3.9) converges as $z \nearrow z_{\max} = p$, and the critical density will be given by the value of the RHS at this point, as this is the largest density for which (3.9) can be solved. As $\langle n_1(z) \rangle$, given by ${}_2\mathrm{F}_1(2,2;2+b;z/p)/(p(1+b){}_2\mathrm{F}_1(1,1;1+b;z/p))$ converges to a finite value, $\langle n_1 z \rangle / L$ will go to zero in the thermodynamic limit and the critical density ρ_c will be given by the second term on the RHS at $z = p$,

$$\rho_c = \frac{p}{1-p} \,. \tag{3.10}$$

For densities above this value the density constraint cannot be solved and it is known (see for example [4]) that a condensation must occur whereby the excess density is taken up by a single site—in this case the defect site—and thus a single site will hold a finite fraction of the total number of particles in the system. Numerical studies indicate that this condensate will appear on the slow site for this system.

For $b < 2$ condensation takes place via mechanism B. The RHS of (3.9) diverges as $z \nearrow z_{\max} = p$, through the divergence of $\langle n_1(z) \rangle$. Thus the density constraint equation (3.9) can be solved for any finite density by taking z arbitrarily close to its maximum allowed value of p. In this mechanism the interpretation of the condensation is clear: the density constraint equation (3.9) is satisfied for any finite density above the critical density by choosing the fugacity z such that $\langle n_1(z) \rangle$ becomes of the order of the system size L and thus this site must hold a finite fraction of all the particles in the system. For this case, the critical density itself is found by taking the limit $z \nearrow p$ in such a way as to ensure $\langle n_1(z) \rangle / L$ tends to zero, i. e. the system is not in the condensed phase. Doing this yields the same critical density as for mechanism A $\rho_c = p/(1-p)$ (3.10).

3.3.2 Grand canonical analysis of the current-density behaviour

The current density behaviour for the SDS system will now be analysed in detail for the two condensation mechanisms. The relation (2.24) equating the fugacity z to the current v is useful here, although it should be noted that this relation is only strictly true in the thermodynamic limit. The convergence properties of the hypergeometric function [29] are also useful

$$\text{if} \quad c > a+b \qquad \lim_{x \to 1^-} {}_2\mathrm{F}_1(a,b;c;x) = \frac{\Gamma(c)\,\Gamma(c-a-b)}{\Gamma(c-a)\,\Gamma(c-b)} \tag{3.11}$$

$$\text{if} \quad c = a+b \qquad \lim_{x \to 1^-} \frac{{}_2\mathrm{F}_1(a,b;c;x)}{|\ln(1-x)|} = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \tag{3.12}$$

$$\text{if} \quad c < a+b \qquad \lim_{x \to 1^-} \frac{{}_2\mathrm{F}_1(a,b;c;x)}{(1-x)^{c-a-b}} = \frac{\Gamma(c)\,\Gamma(a+b-c)}{\Gamma(a)\,\Gamma(b)} \tag{3.13}$$

where $\Gamma(x)$ is the usual Gamma function [29].

For $b > 2$, the system will condense via mechanism A and using (3.11) the average number of particles on the defect site will converge to the finite value

$$\lim_{z \nearrow p} \langle n_1(z) \rangle = \frac{1}{b-2} , \tag{3.14}$$

as $z \nearrow z_{\max} = p$. Thus the density constraint equation (3.9) cannot be satisfied for $z \leq p$, when $\rho > \rho_c$ and $z > p$ must be considered. This implies an overshoot in the grand-canonical current. Note that the fact that the average number of particles converges to a finite value means there is no condensate on the defect site for $z \nearrow p$. This is because a condensate comprises of a finite fraction of the total number of particles in the system, so the average number of particles on the site must diverge with the system size. Indeed for $b > 3$ the average number of particles on the defect site can be less than one, which clearly does not indicate the presence of a condensate. The overshoot is investigated by imposing particle number dependent cut-offs in the sums that make up $\langle n_1(z) \rangle$. These sums are cut-off at $n_1 = N$, thus restricting the slow site never to have more than the total number of particles in the system.

$$\langle n_1(z) \rangle = \frac{z}{p(1+b)} \frac{{}_2\mathrm{F}_1^{(N)}(2,2;2+b;z/p)}{{}_2\mathrm{F}_1^{(N)}(1,1;1+b;z/p)} , \tag{3.15}$$

where

$${}_2\mathrm{F}_1^{(N)}(a,b;c;x) = \sum_{m=0}^{N} \frac{(a)_m (b)_m}{(c)_m} \frac{x^m}{m!} , \tag{3.16}$$

i. e. this is a truncated hypergeometric sum.

The asymptotics of z are now determined. The following assumption for the form of z/p

$$\frac{z}{p} = \exp(\alpha(N)) , \tag{3.17}$$

is made, with $\alpha(N)$ being a small N dependent quantity. In order that the density constraint equation (3.9) is satisfied, $\langle n_1(z) \rangle = \mathcal{O}(N)$ is required. As z is taken above p, this will first be true when the numerator of (3.15) is $\mathcal{O}(N)$ and the denominator $\mathcal{O}(1)$. The value that $\langle n_1(z) \rangle$ takes for $z > p$ is evaluated by replacing the sums of (3.15) by integrals over asymptotic forms of the summands

$${}_2\mathrm{F}_1^{(N)} = \sum_{m=0}^{N} \frac{(a)_m (b)_m}{(c)_m} \frac{x^m}{m!} \simeq A + \int_1^N \mathrm{d}k \, \frac{\exp(k\alpha(N))}{k^b} , \tag{3.18}$$

where A is a finite error term of order 1. Thus $\langle n_1(z) \rangle$ may be approximated as

$$\langle n_1(z) \rangle \simeq \frac{G_{b-1}}{A + G_b} , \tag{3.19}$$

where

$$G_b = \int_1^N \mathrm{d}k \, \frac{\exp(k\alpha(N))}{k^b} \,, \tag{3.20}$$

and the finite error term has been dropped from the numerator as the integral in the numerator is expected to be $\mathcal{O}(N)$. The asymptotic behaviour of G_b can be found by integrating by parts giving

$$G_b \sim \frac{\exp(N\alpha(N))}{N^b \alpha(N)} \quad \text{for} \quad N\alpha(N) \gg 1 \,. \tag{3.21}$$

Thus it is clear that $G_b = \mathcal{O}(1)$ guarantees $G_{b-1} = \mathcal{O}(N)$ if $N\alpha(N) \gg 2$. Therefore $\langle n_1(z) \rangle = \mathcal{O}(N)$ and the density equation (3.9) will be satisfied for large-but-finite systems when

$$\frac{\exp(N\alpha(N))}{N^b \alpha(N)} = \mathcal{O}(1) \,. \tag{3.22}$$

This implies

$$\alpha(N) \simeq (b-1)\frac{\ln N}{N} + \frac{\ln(\ln N)}{N} + \mathcal{O}\left(\frac{1}{N}\right) \,, \tag{3.23}$$

which can be seen by solving (3.22) iteratively. Thus the grand-canonical analysis predicts an overshoot when $b > 2$. Recalling that in the grand canonical ensemble $v = z$, the current is predicted to behave as

$$v \simeq p \exp\left[(b-1)\frac{\ln N}{N} + \frac{\ln(\ln N)}{N} + \mathcal{O}\left(\frac{1}{N}\right)\right] \,, \tag{3.24}$$

for particle numbers above the critical particle number, given by $L\rho_c$.

For $b \leq 2$, the system will condense via mechanism B. Here the average number of particles on the defect site will diverge when $z \nearrow z_{\max} = p$. Thus the density constraint equation (3.9) can be solved by taking z arbitrarily close to p. Thus one need not take $z > p$ and no overshoot is predicted by the grand-canonical analysis in this case. The large-but-finite system is now analysed by determining how close z must be taken to p, in terms of the system size L, to solve (3.9). Using the convergence properties of the hypergeometric functions (3.11 – 3.13), the behaviour of $\langle n_1 \rangle$ as $z \nearrow p$ can be analysed and it is found that

$$\lim_{z \nearrow p} \frac{\langle n_1(z) \rangle}{|\ln(1 - z/p)|} = 1 \qquad \text{for} \quad b = 2 \tag{3.25}$$

$$\lim_{z \nearrow p} \langle n_1(z) \rangle \, (1 - z/p)^{2-b} = \frac{(b-1)^2 \pi}{\sin(\pi(b-1))} \qquad \text{for} \quad 1 < b < 2 \tag{3.26}$$

$$\lim_{z \nearrow p} \langle n_1(z) \rangle \, (1 - z/p) \, |\ln(1 - z/p)| = 1 \qquad \text{for} \quad b = 1 \tag{3.27}$$

$$\lim_{z \nearrow p} \langle n_1(z) \rangle \, (1 - z/p) = (1 - b) \qquad \text{for} \quad b < 1 \,, \tag{3.28}$$

where the identity

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin \pi x} , \tag{3.29}$$

has been used to obtain (3.26). Thus, for $b \leq 2$, the density constraint equation (3.9) can be satisfied by making the choice

$$1 - z/p \sim \exp[-L(\rho - \rho_c)] \qquad \text{for} \quad b = 2 \tag{3.30}$$

$$1 - z/p \sim \left[\frac{(b-1)^2 \pi}{\sin(\pi(b-1))(\rho - \rho_c)} \frac{1}{L} \right]^{1/(2-b)} \qquad \text{for} \quad 1 < b < 2 \tag{3.31}$$

$$1 - z/p \sim \frac{1}{(\rho - \rho_c)} \frac{1}{L \ln L} \qquad \text{for} \quad b = 1 \tag{3.32}$$

$$1 - z/p \sim \frac{(1-b)}{(\rho - \rho_c)} \frac{1}{L} \qquad \text{for} \quad b < 1 . \tag{3.33}$$

Thus for $b < 2$ the grand-canonical analysis predicts that the particle current approaches its expected saturation value from below for large-but-finite systems.

An advantage for the SDS system is that the current below the critical density is straightforward to calculate. Noting that in the un-condensed phase the first term on the RHS of (3.9) goes to zero, the equation can simply be inverted to give

$$z = \rho/(1+\rho) . \tag{3.34}$$

The equivalence of the current and the fugacity, z, is then invoked. This equivalence and equation (3.34) are strictly only true in the thermodynamic limit, but as can be seen from Figure 3.1, the finite size simulations agree extremely well with this thermodynamic prediction.

3.3.3 Simulations and exact numerics in the canonical ensemble

Numerical analysis within the canonical ensemble can be used to investigate the behaviour of the SDS system and go some way to testing the theory put forward in the previous section. As detailed in Sections 2.6.1 and 2.6.2, the ZRP can be simulated with dynamical Monte-Carlo algorithms and the normalisation (2.7)—the equivalent of the canonical partition function for equilibrium systems—can be evaluated numerically.

The current v, as calculated from recursive numerical evaluation of the normalisation $\mathrm{Z}(N, L)$ using relation (2.45), displays the overshoot very clearly, see Figures 3.3 and 3.4.

In Figure 3.3 it is clearly seen than an overshoot is present for $b > 0$. This is in contrast to the grand canonical prediction that an overshoot should only occur for

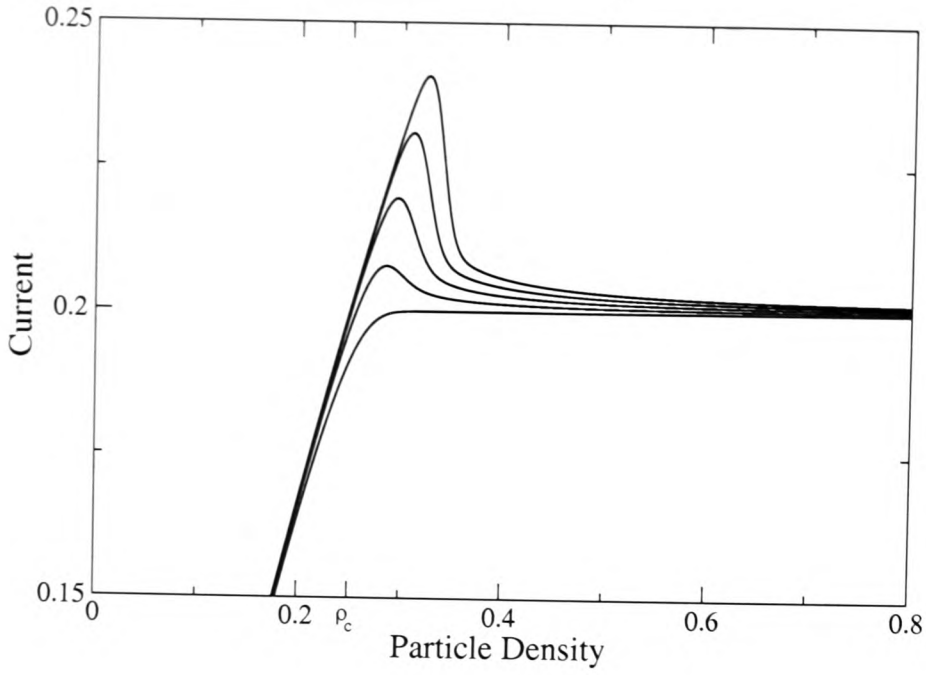


Figure 3.3: Current density diagram as calculated from exact numerical evaluation of the partition function for the SDS system with $L = 1000$, $p = 0.2$ and $b = 0, 1, 2, 3, 4$ from bottom to top. The critical density for the systems in the thermodynamic limit, ρ_c, is equal to 0.25 (independent of the value of b) and is marked on the graph.

$b > 2$ in the SDS system. It can also clearly be seen that the severity of the overshoot is increasing for increasing b.

The results presented in Figure 3.4 verify that the overshoot is a finite size effect as the severity of the overshoot is seen to decrease with increasing system size. It is also apparent that the system is heading towards the expected thermodynamic behaviour as the system size is increased.

The results from numerical evaluation of the partition function also agree closely with those taken from dynamical Monte-carlo simulations as one would expect. In Figure 3.4 the current output from numerical evaluation of the normalisation is compared with that measured from a simulation. The results agree very well, even around the expected critical density, where one might not expect simulations to pick up the true behaviour due to, for example, large fluctuations which are often associated with transition points.

As the simulations appeared to ably reproduce the behaviour of the system around the critical point, they are used to investigate what is actually happening in the overshoot region. Thus far the existence of the overshoot has been confirmed and its general form analysed but little has been said about what behaviour it corresponds to in the system other than in terms of the particle current v. As was detailed in the introduction the overshoot is a phenomenon that often occurs in the presence conden-

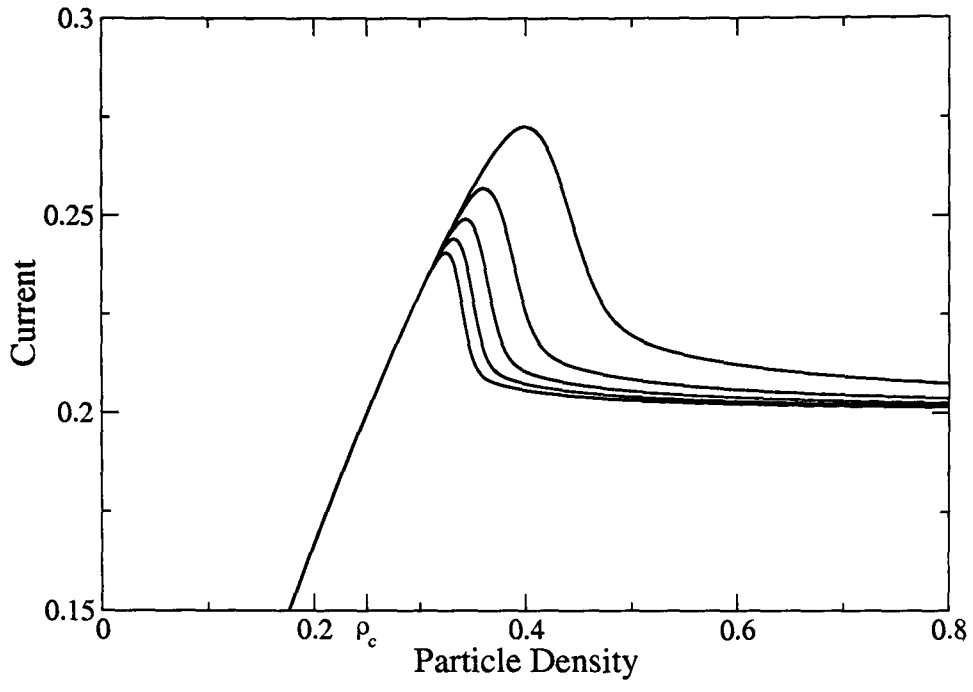


Figure 3.4: Current density diagram as calculated from exact numerical evaluation of the partition function for SDS system with $p = 0.2$, $b = 4$ and $L = 200, 400, 600, 800, 1000$ from top to bottom. The critical density for the system in the thermodynamic limit, ρ_c, is equal to 0.25 and is marked on the graph.

sation transitions. Simulations are used to probe the condensation-like behaviour in large-but-finite systems and identify if this is linked to the overshoot behaviour.

The current-density diagram can be divided into three distinct regions:

- $\boldsymbol{\rho < \rho_c}$—Below the critical density, the large-but-finite system behaviour and the expected thermodynamic behaviour match quite closely as can be seen in Figure 3.1. The system is expected to be in an un-condensed, fluid phase where the particles are evenly spread throughout the system. This is the low density region.

- $\boldsymbol{\rho \gg \rho_c}$—Far above the critical density, the large-but-finite system behaviour and the expected thermodynamic behaviour again match up quite well as can be seen in Figure 3.1. The system is expected to be in a condensed phase with a single site holding a finite fraction of the total number of particles in the system. This is the high density region.

- $\boldsymbol{\rho \gtrsim \rho_c}$—Above, but close to, the critical density the large-but-finite system shows non-monotonic behaviour in the current, whereas the expected thermodynamic behaviour is a constant. This is the overshoot region and the behaviour of the configuration of a large-but-finite system is to be determined in this region.

Monte-carlo simulations were used to investigate the behaviour in each of these three

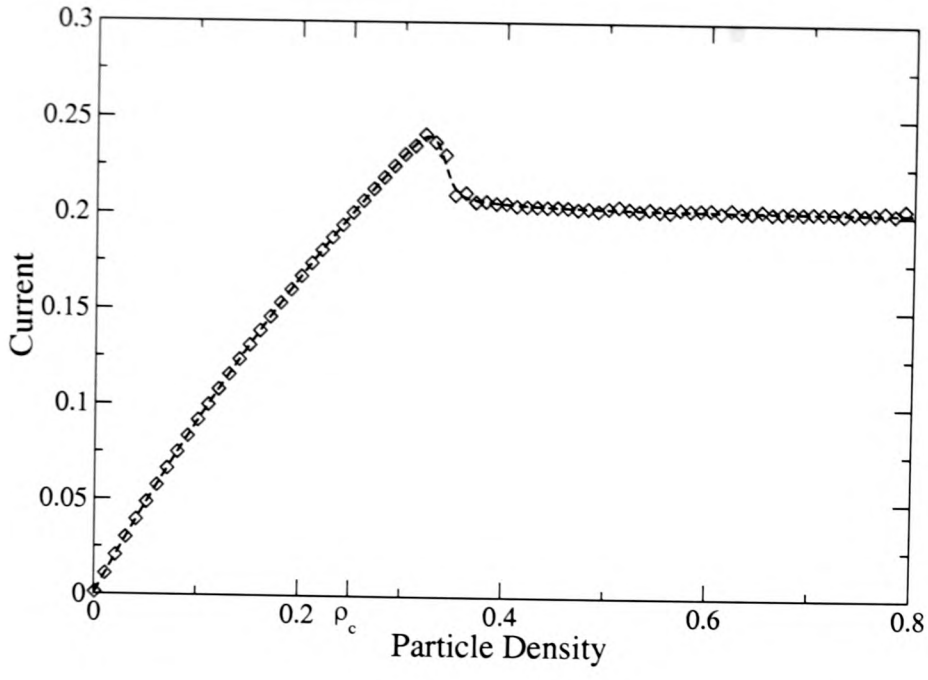


Figure 3.5: Comparison of the current-density diagram for the SDS system with $L = 1000$, $p = 0.2$ and $b = 4$, as calculated from exact numerical calculation of the partition function and Monte Carlo simulation with direct measurement of the particle current. Simulations were run from random initial conditions for 10^6 Monte Carlo sweeps at density intervals of 0.01. The critical density for the system in the thermodynamic limit, ρ_c, is equal to 0.25 and is marked on the graph.

regions, in particular that in the overshoot region.

For the SDS system, there is a single defect site which will have a slower hop rate for a large enough occupation than any of the other sites. Thus it is expected that a condensate will nucleate and be stable on this defect site only. Thus a useful measurement that can be taken during a simulation is the time-series of the number of particles on the defect site. The fraction of the particles on the slow site can be taken to be an order parameter for the system.

The behaviour in the low density and high density regions is straightforwardly confirmed as can be seen in Figure 3.6. The number of particles on the slow-site throughout the time-series taken at $\rho = 0.23$, i. e. the low density phase has an average value close to the value if the particles were evenly spread over the sites, although it is a little bit bigger. Similarly the number of particles on the slow-site throughout the time-series taken at $\rho = 0.5$, i. e. the high density phase, has an average value that is close to $L(\rho - \rho_c)$, the excess number of particles above the critical number. This is the number of particles one would expect to see on a condensed site in the condensed phase.

The behaviour within the overshoot region itself is now discussed. As can be seen

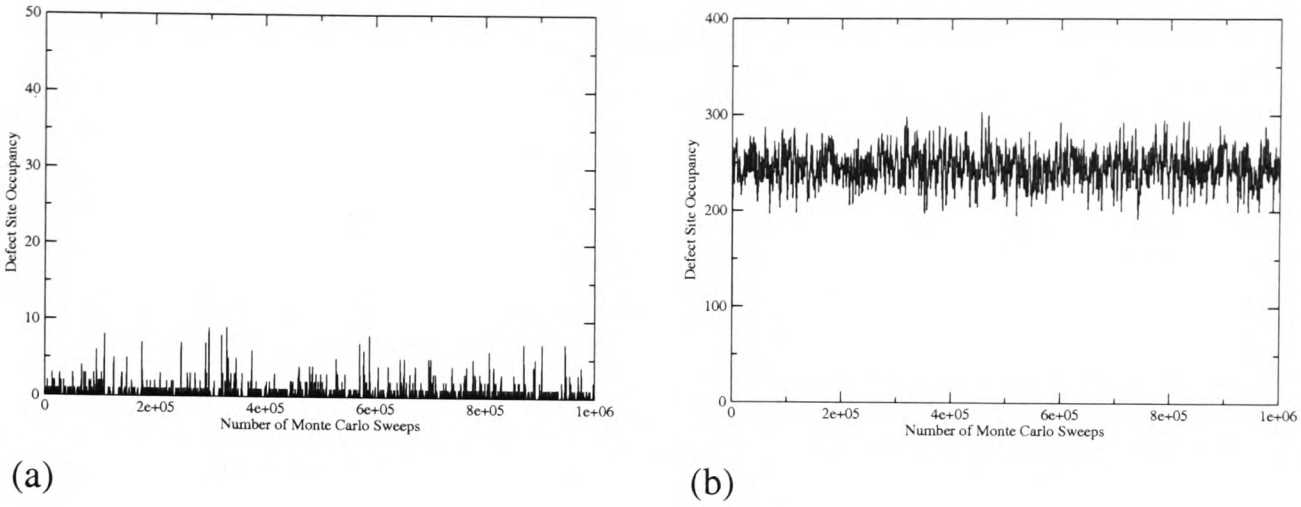


Figure 3.6: Time series of the occupation of the defect site from simulations of the SDS system with $L = 1000$, $p = 0.2$ and $b = 4$. Simulations were run for 10^6 Monte Carlo sweeps from random initial conditions. (a) Simulation run at particle density $\rho = 0.23$, which is well within the fluid phase. (b) Simulation run at particle density $\rho = 0.5$, which is well within the condensed phase.

in Figure 3.5 in this region the current is initially increasing, reaches a maximum, drops and tends towards a constant value with increasing density. As previously observed, below the critical density in the fluid phase the finite system behaviour closely matches that expected of the thermodynamic system and the current is monotonically increasing. The current in the finite system seems to continue increasing on an extrapolation of the fluid phase line to a density above the critical density. Thus a possibility for the first part of the overshoot region is a continuation of the fluid phase to higher densities than would be possible for the thermodynamic system. This interpretation seems to be correct as can be seen in Figure 3.7 (a), where the occupation of the slow site is seen to fluctuate around a low average value as it would in the fluid phase. Then at higher densities the current begins to decrease; as can be seen from Figure 3.7 (b) this appears to coincide with an unstable putative condensate beginning to form on the defect site. This can be thought of as a temporal coexistence between the fluid and condensed phases: the system spends the majority of its time in one phase or the other and moving between the two takes relatively little time. At even higher densities the current begins to get close to the expected thermodynamic saturation value and the gradient shallows. This seems to coincide with the condensate beginning to become quite stable as seen in Figure 3.7 (c).

Thus, from numerical evaluation of the normalisation $\mathrm{Z}(N, L)$ (Figure 3.5) and dynamical Monte-carlo simulation (Figure 3.7), the overshoot region is interpreted as follows: Initially it is an extension of the fluid phase to densities above the expected

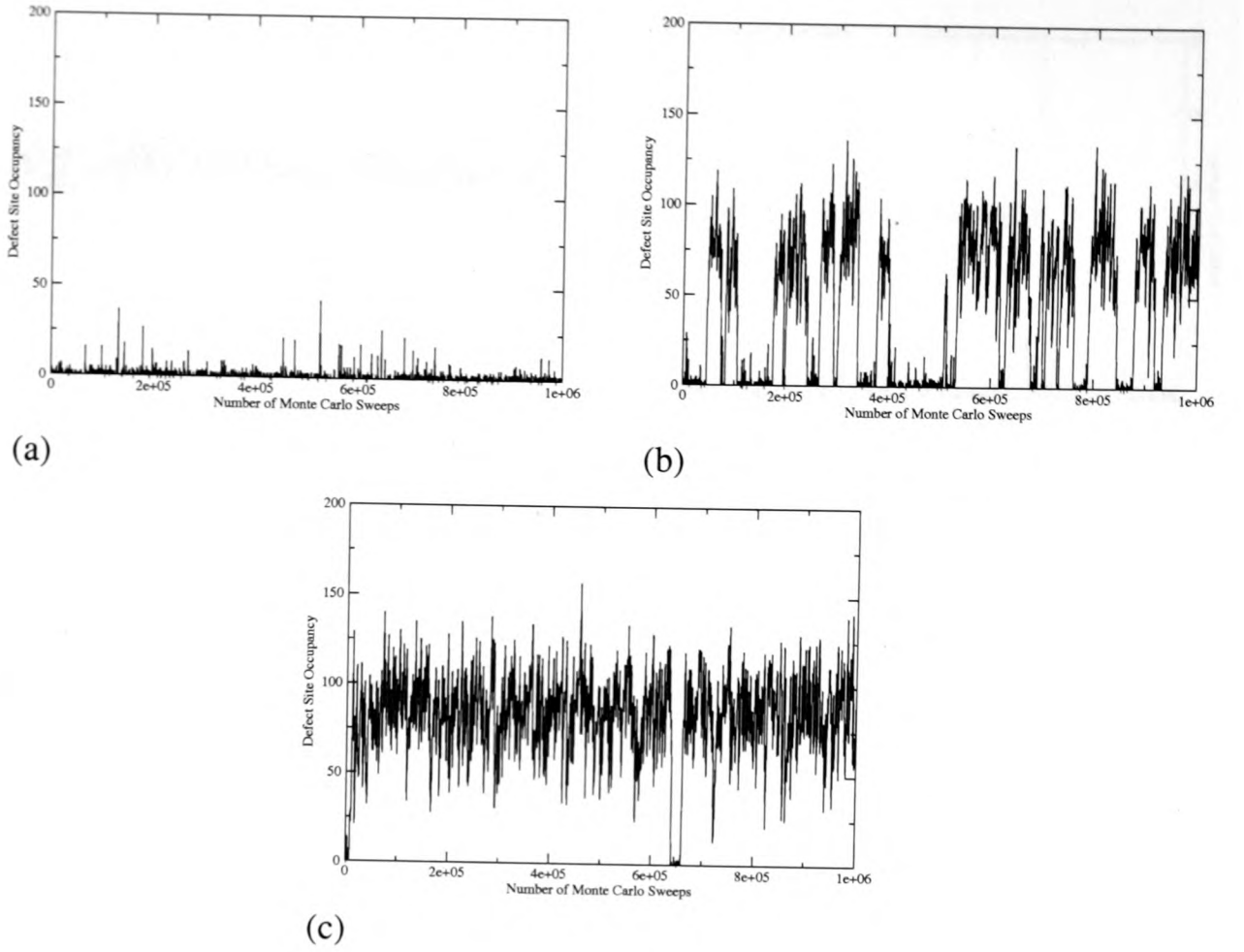


Figure 3.7: Time series of the occupation of the defect site from simulations of the SDS system with $L = 1000$, $p = 0.2$ and $b = 4$. Simulations were run for 10^6 Monte Carlo sweeps from random initial conditions. (a) Simulation run at particle density $\rho = 0.29$, which is within the region of the overshoot that appears to be a continuation of the fluid-phase line. (b) Simulation run at particle density $\rho = 0.34$, which is within the region of the overshoot where the current is decreasing rapidly. (c) Simulation run at particle density $\rho = 0.35$, which is within the region of the overshoot where the current is close to the saturation value.

thermodynamic critical density, this region extends from ρ_c to the maximum of the overshoot; it then gives way to a 'coexistence' region where a putative condensate begins to form on the defect site, but is unstable and the system oscillates between fluid and condensed states, this region exists where the current is decreasing most steeply; finally the condensate begins to become stable, this region is where the current begins to decrease less steeply and tend towards the expected thermodynamic saturation value.

Thus far only numerical evidence has been given for the above interpretation of the overshoot region. A grand-canonical analysis has been presented which does not correctly predict the parameters which allow an overshoot. However, a canonical

analysis is also possible for the SDS system and this also supports the interpretation of the overshoot region. This canonical analysis is presented in the next section.

3.3.4 Canonical analysis

A useful property of the SDS system is that it has a particularly simple form for the normalisation $\mathrm{Z}(N, L)$. Due to all the non-defect sites having the same constant hop rates the non-defect sites can be summed out of the normalisation easily. Recalling the general form of the canonical normalisation (2.7) and inserting the $f_\mu(n)$ functions (3.2) the following is found

$$\mathrm{Z}(N, L) = \sum_{n_1} p^{-n} \frac{n!}{(1+b)_n} \sum_{n_2} 1 \sum_{n_3} 1 \cdots \sum_{n_L} 1 \, \delta\left(\sum_{\mu} n_\mu - L\right) . \tag{3.35}$$

Performing the sums over $n_2, \ldots, n_L$ simply gives a combinatoric factor from the delta function: the contribution from the non-defect sites is simply the number of ways of arranging the particles that are not on the defect site. The normalisation thus reduces to a sum over the occupation of the defect site

$$\mathrm{Z}(N, L) = \sum_{n=0}^{N} p^{-n} \frac{n!}{(1+b)_n} \binom{N+L-n-2}{L-2} . \tag{3.36}$$

Note that for simplicity n_1 has been re-labelled as n.

The normalisation $\mathrm{Z}(N, L)$ has thus been reduced to a single sum over the unnormalised probabilities that the defect site has n particles on it, i. e. the n^{th} term of the sum is proportional to the probability that the defect site has n particles on it in the steady state. Thus studying which terms dominate the sum gives an indication of what state the system will be in. If $\mathrm{Z}(N, L)$ is dominated by terms at low n, the corresponding system will be in the fluid phase and if it is dominated by terms at high n it will be in the condensed phase. If there are two dominating parts of the sum, at low and high n, then this is interpreted as some kind of phase coexistence. Given the results found in the numerical study of the system, investigating the maxima of the normalisation provides a broad explanation of the overshoot and the behaviour seen in the system.

The turning points in a sum are found by looking for consecutive terms that have a ratio tending to 1. Solving for these points in the SDS system gives a quadratic equation in n, the solutions of which are:

$$\begin{aligned} n = {} & \frac{L}{2}[\rho - \rho_c] + A/2 \\ & \pm \frac{L}{2}\left[(\rho-\rho_c)^2 + \frac{1}{L}\left\{2(\rho-\rho_c)A - 4b\rho_c(1+\rho)\right\} + \frac{1}{L^2}\left\{A^2 + 4b\rho_c\right\}\right]^{\frac{1}{2}} \end{aligned} \tag{3.37}$$

where $\rho_c = p/(1-p)$ is the thermodynamic critical density and the constant $A = \rho_c/p + \rho_c(1+b)$ has been introduced to lighten the equation.

Now, the number of turning points in the physical region $0 < n < N, n \in \mathbb{R}$ is given by the number of solutions of (3.37) in this region. The nature of the turning points can then be found by considering whether the ratios of the first two terms of the sum and the last two terms of the sum are greater than or less than 1. For this case the ratio of the final two terms in the sum indicates that the terms are always decreasing at this boundary.

The behaviour of the system is most clearly revealed by considering the sequence of boundary and turning points as the density of the system is increased from zero. First, note that the roots of (3.37) first become real at a density ρ_2 given by

$$\rho_2 = \frac{p}{1-p} - \frac{1}{L}\left(\frac{1+p(1-b)}{1-p}\right) + \left[\frac{1}{L}\frac{4bp}{(1-p)^2} - \frac{1}{L^2}\frac{8bp}{(1-p)^2}\right]^{\frac{1}{2}}, \tag{3.38}$$

and that this is always larger than $\rho_c = p/(1-p)$. The sequence of turning points is best displayed in the profiles of the logarithm of the terms in the sum. As shown in Figure 3.8, they go through the sequence: (a) For $\rho < \rho_2$ there are no turning points, only a boundary maximum at $n = 0$; (b) at $\rho = \rho_2$ a stationary point emerges and the boundary maximum remains; (c) for $\rho > \rho_2$ the stationary point splits into a local minimum and a local maximum with the maximum at higher n; (d) as the density is increased further the minimum moves to smaller n until it hits the boundary and this becomes a boundary minimum. For this full sequence it has been assumed that $p < 1/(1+b)$; for $p \geq 1/(1+b)$ the low n boundary never becomes a minimum. However, this makes little difference to the behaviour of the system and none to the interpretation of the overshoot.

The sequence is now described in more detail and related to the behaviour of the system. At low density, $\rho < \rho_2$, there are no turning points in the region of interest. Thus the lowest terms in the sum dominate, with the zeroth term the largest. This corresponds to the fluid phase: the defect site has low occupancy with the highest probability. Thus the fluid phase continues to a density ρ_2, greater than the expected critical density. At densities greater than ρ_2, a maximum emerges at high n, but can be of similar magnitude to the boundary maximum at $n = 0$. Thus the defect site is likely to have both few and many particles on it; this is interpreted as a coexistence of phases and the system would be expected to move between these phases as seen in the simulations, see Figure 3.7 (b). At higher densities $\rho \gg \rho_2$, the maximum at large n dominates over the boundary and so the defect site will have a large number of particles on it and the system will be in the condensed phase.

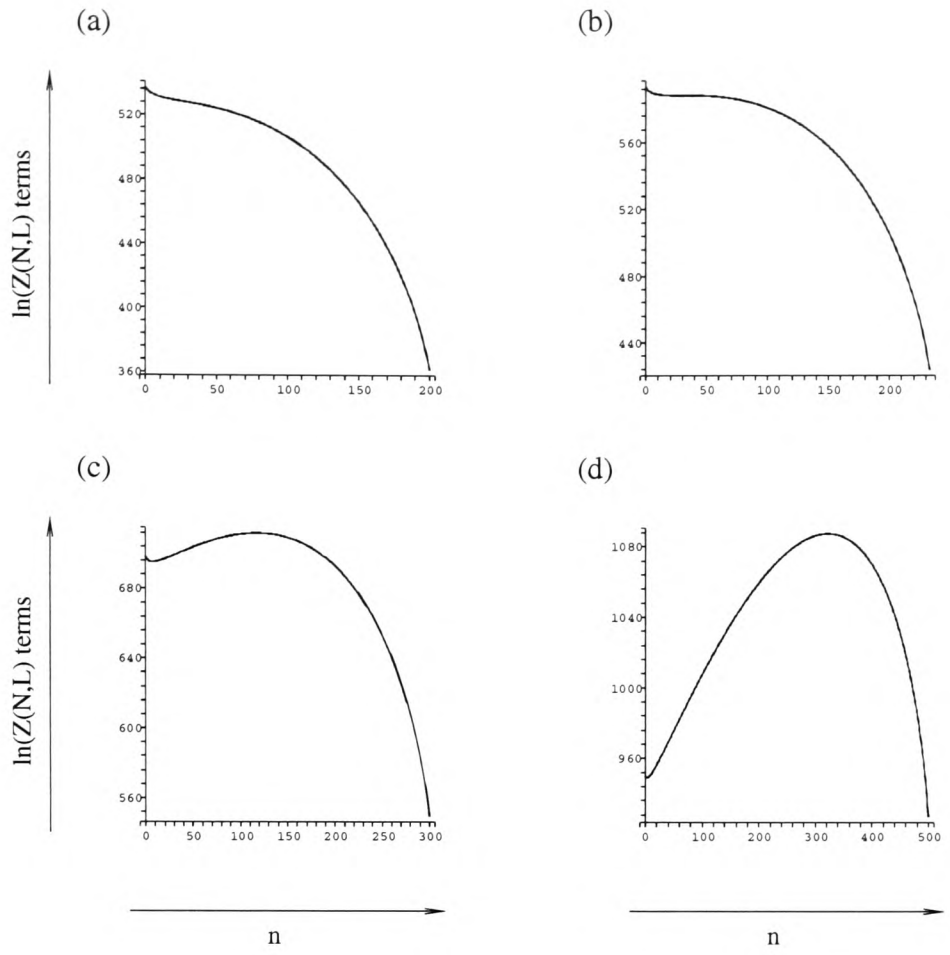


Figure 3.8: Series showing the logarithms of the terms from the normalisation sum $Z(N, L)$ (3.36), for a system of size $L = 1000$ with $b = 4$ and $p = 0.15$ at the following particle numbers. (a) $N = 200$ ($\rho < \rho_2$); here the sum has a boundary maximum at $n = 0$ and no turning points. (b) $N = 233$ ($\rho = \rho_2$); here there is a boundary maximum at $n = 0$ and a stationary point at higher n. (c) $N = 300$ ($\rho > \rho_2$); here there is a boundary maximum at $n = 0$ and a greater maximum at high n. (d) $N = 500$ ($\rho \gg \rho_2$); here there is a boundary minimum at $n = 0$ and a maximum at high n.

Now, the ratio of the final two terms of the sum (3.35) indicates that the terms are always decreasing at the upper end of the sum and the ratio of the first two terms indicates that for small enough densities the terms will also be decreasing at the low end. Thus the normalisation sum can only have a maximum at high n, and hence a condensate, when real roots of (3.37) emerge. This happens at density ρ_2 which is greater than the thermodynamic critical density ρ_c. Note that $\lim_{L\to\infty} \rho_2 = \rho_c$; thus for a thermodynamic system a maximum will emerge at high n and a condensate will be present for densities $\rho > \rho_c$. Hence the finite system canonical analysis recovers the expected thermodynamic behaviour in the appropriate limit. For large-but-finite systems condensation cannot occur until $\rho > \rho_2$, thus the fluid phase continues to a density greater than ρ_c. For finite systems ρ_2 is greater than the critical density

$\rho_c = p/(1-p)$ by an amount $\mathcal{O}(L^{-1/2})$:

$$\rho_2 = \frac{p}{1-p} + \frac{2}{1-p}\left(\frac{bp}{L}\right)^{1/2} + \mathcal{O}\left(\frac{1}{L}\right) . \tag{3.39}$$

As ρ_2 is the first point at which a high n maximum emerges, this density is expected to coincide with the peak of the overshoot as this is where the coexistence region is assumed to begin. Comparing the peak of the overshoot, measured from numerical evaluation of the normalisation, to ρ_2, as calculated from the canonical analysis of the normalisation, yields good agreement for $b \approx 1 \ldots 6$, as seen in Figure 3.9. Outwith this range it is believed that further finite size effects are becoming important.

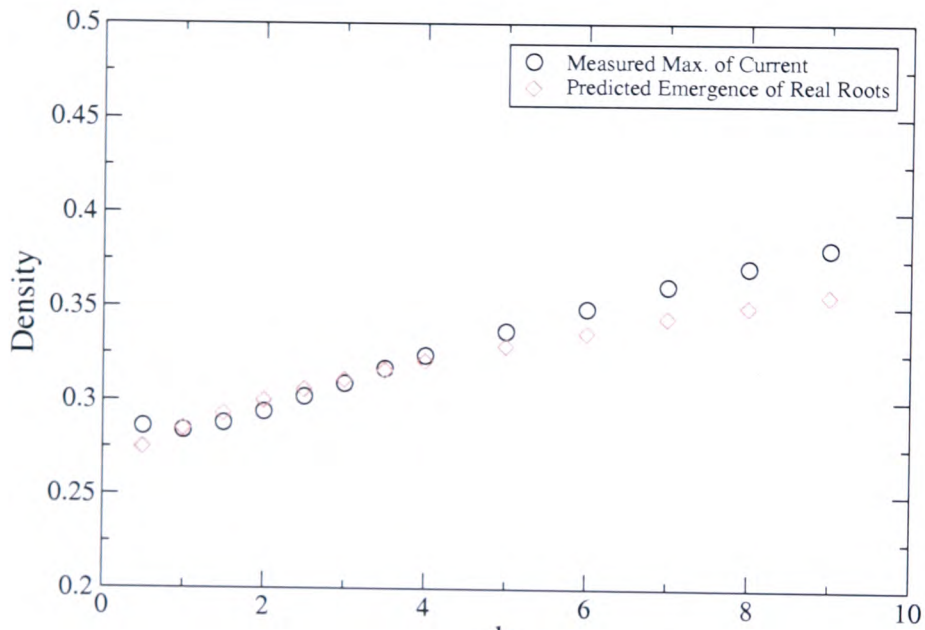


Figure 3.9: Comparison between the maximum of the current as measured from exact numerical calculation and the predicted emergence of real roots from the theory for an SDS system with $L = 1000$, $p = 0.15$ and b varying between 0.5 and 9.

3.3.5 Canonical calculation of the current

Another convenience of the single sum form of the normalisation for the SDS system (3.35) is that the asymptotic expansion of this can be found. This is achieved by first asymptotically expanding the summand for large n, then approximating the sum itself by an integral and asymptotically expanding this integral. The relation (2.9) can then be used to give the asymptotics of the particle current and thus further description of the overshoot from a canonical viewpoint. See Appendix B for the details of how to do this.

The asymptotic form for the current is found to be

$$v = p\left[1 + \frac{1}{L}\frac{b}{\rho - \rho_c} + \mathcal{O}\left(\frac{1}{L^2}\right)\right] \tag{3.40}$$

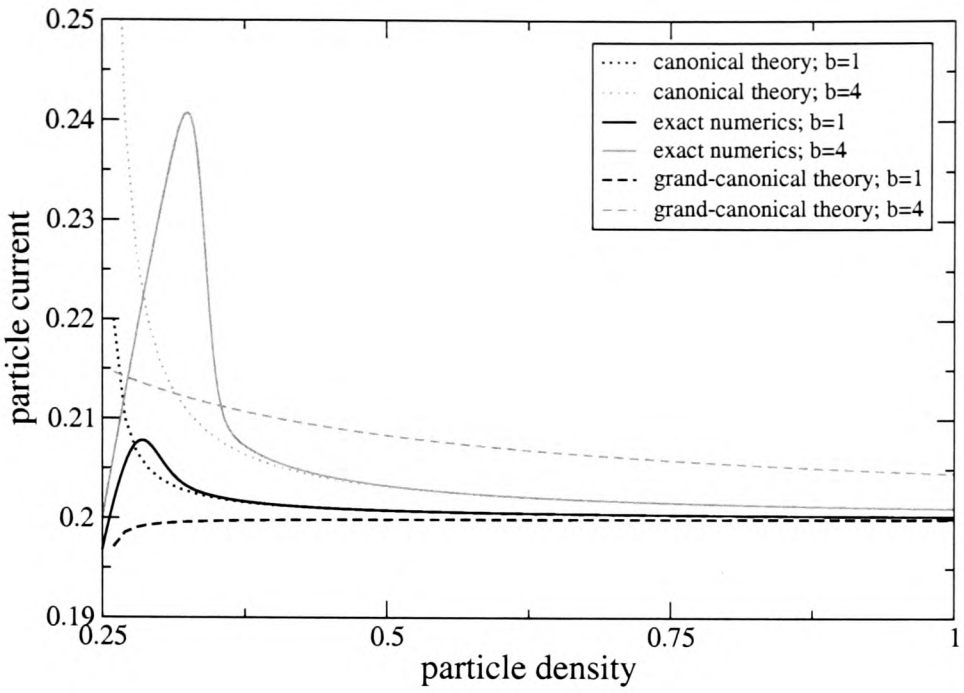


Figure 3.10: Comparison between the theory prediction for the overshoot from both the canonical (dotted lines) and grand-canonical (dashed lines) approaches and the overshoot as calculated by exact numerical evaluation of the partition function (solid lines). Data taken from the SDS system defined by (3.1) with $b = 1$ (black lines) and $b = 4$ (grey lines). As expected, the canonical theory matches the exact calculation at high density where the theory is valid. The grand-canonical theory line for $b = 1$ is completely different from the canonical lines due to the failure of the grand-canonical analysis to predict an overshoot for $b \leq 2$. The gradient of the grand-canonical theory line for $b = 4$ differs by a factor $\sim \ln N$ to leading order.

This repeats the canonical prediction that the overshoot should be present for $b > 0$. It should be noted that this expansion is only valid for large n; specifically, it is only valid for densities appreciably greater than the critical density, such that $\rho - \rho_c \gg b/L$. This result is compared with the results from the numerical computation of $Z(N, L)$ and the grand canonical analysis in Figure 3.10. It is found that this disagrees with the grand canonical expression —by a factor $\sim \ln N$ to leading order for $b > 2$— but agrees with the numerical results, from simulations in the canonical ensemble, quite closely.

3.4 Two defect sites

With the overshoot phenomenon now studied and to some extent understood, a natural question that arises is: Can any interesting interplay effects be seen if extra defect sites are added to the SDS system? The answer is yes.

The simplest case to consider is an SDS-like system with two defect sites instead of one. These are chosen to be sites 1 and 2, for simplicity, and are allowed to have different rates. A simple choice for the hop rates that allows for interesting behaviour is

$$u_1(n) = \theta(n)p_1(1 + b_1/n) \tag{3.41a}$$

$$u_2(n) = \theta(n)p_2(1 + b_2/n) \tag{3.41b}$$

$$u_{\mu\neq 1,2}(n) = \theta(n)\ . \tag{3.41c}$$

A careful choice of the parameters p_1, b_1, p_2, b_2 can reveal some interesting behaviour. For example, making $p_2 < p_1$ results in site 2 presenting the slowest possible option for the system and one would expect that in the thermodynamic limit any condensate would form on this site. This is borne out by the expected thermodynamic critical densities for SDS systems with either of the slow sites removed. However, if b_1, b_2 are chosen such that in an SDS system the overshoot for site 2 reaches to higher density than that for site 1, then it might be expected that a putative condensate would form first on site 1 and then move to site 2 at higher densities. In fact, this is exactly what can be observed in both simulations and average site occupations computed from numerical evaluation of $\mathrm{Z}(N, L)$. The average occupations of sites 1 and 2, as calculated from numerical evaluation of $\mathrm{Z}(N, L)$, are shown in Figure 3.11, where it can clearly be seen that a condensate initially forms on site 1 before moving to site 2 at higher densities.

For this system, the normalisation has reduced to a double sum, over the occupancies of the two defect sites

$$\mathrm{Z}(N, L) = \sum_{n_1=0}^{N} \sum_{n_2=0}^{N-n_1} p_1^{-n_1} p_2^{-n_2} \frac{n_1!}{(1+b_1)_{n_1}} \frac{n_2!}{(1+b_2)_{n_2}} \binom{N+L-n_1-n_2-3}{L-3}\ . \tag{3.42}$$

The story is much the same as for the SDS system: This sum goes from having a dominant peak at $n_1, n_2 \approx 0$, to having a dominant peak at $n_1 \approx 0$ with n_2 large, to a dominant peak at $n_2 \approx 0$ with n_1 large with increasing density. This sequence can be seen in plots of the logarithm of the profile of the terms in the double sum, as shown in Figure 3.12. The interpretation of this is initially the system will be in a state with both defect sites lowly occupied, then the occupation of site 1 will increase while site 2 will remain lowly occupied, then sites 1 and 2 will both have sizeable occupations, finally site 2 will take over and site 1 will have low occupation. This is exactly the sequence seen in Figure 3.11.

Thus with two defect sites further interesting behaviour has been added to the system. It is expected that similar things could be seen for three or more defect

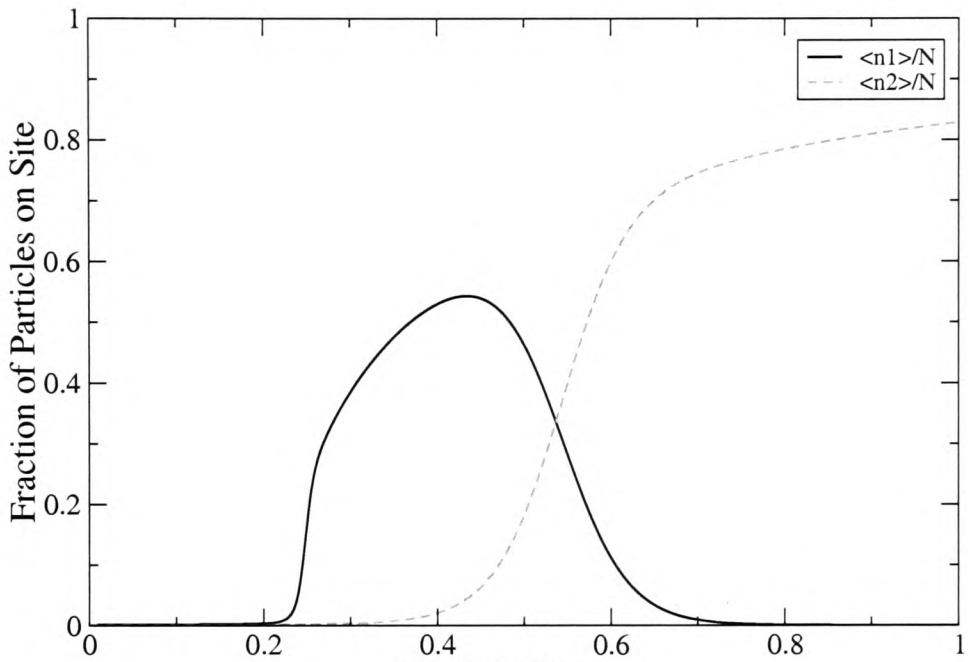


Figure 3.11: Average fraction of the total number of particles on a site versus the global particle density for the two defect site system defined through (3.41). The chosen parameters are $b_1 = 4$, $p_1 = 0.15$, $b_2 = 7$ and $p_2 = 0.145$, meaning that the second site is the preferred site for a condensate in the thermodynamic limit but that the finite size system prefers the first site for low densities. Data were calculated by exact numerical evaluation of the partition function, from which n_1/N (solid black line) and n_2/N (dashed grey line) can be straightforwardly extracted.

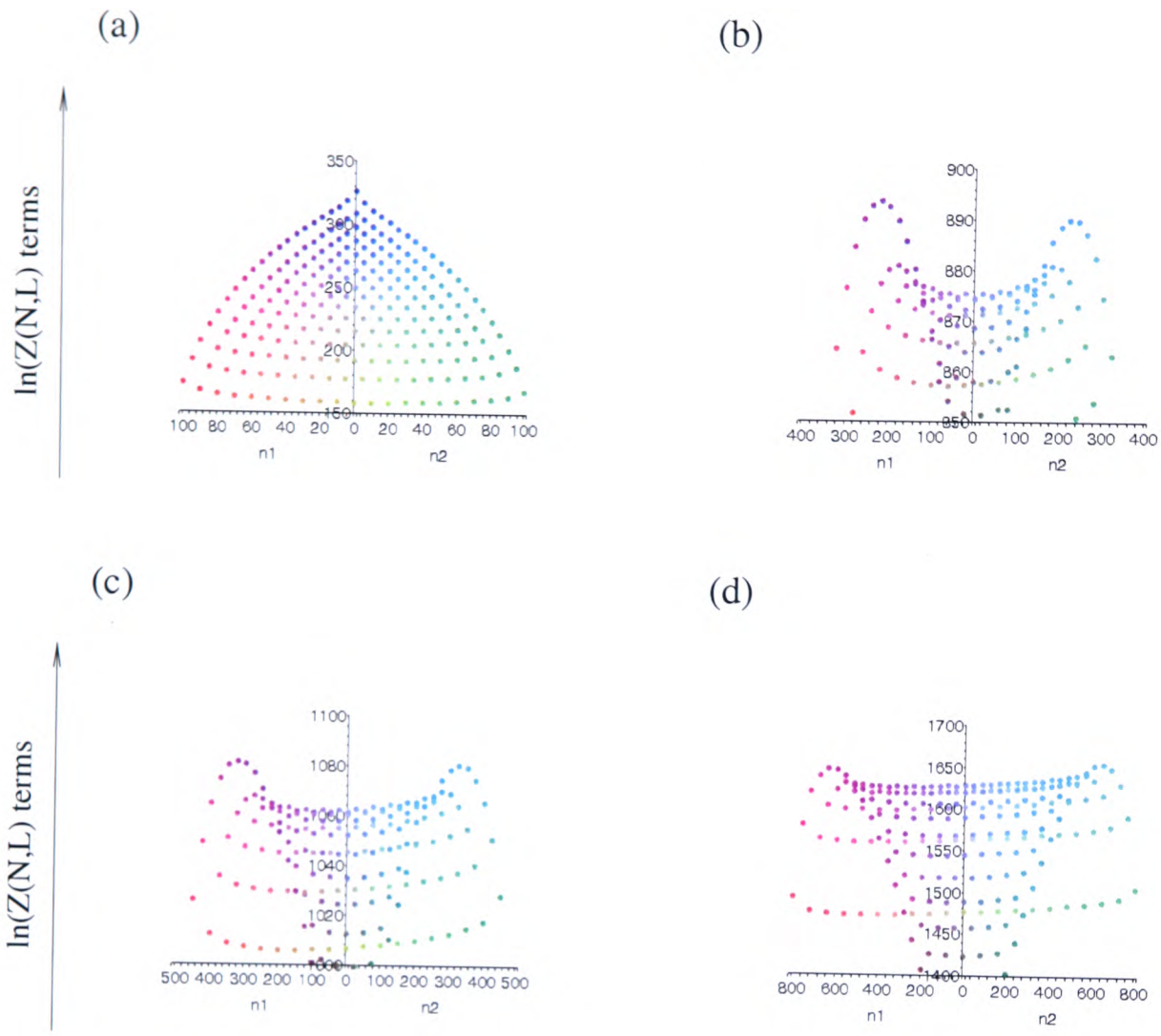


Figure 3.12: Series showing the logarithms of the terms from the normalisation $Z(N, L)$ double sum over n_1 and n_2 for the 2 defect site system (3.42). These were computed for a system of size $L = 1000$ with $b_1 = 4$, $p_1 = 0.15$, $b_2 = 7$ and $p_2 = 0.145$ for varying global particle density. All are shown with the $n_1 = n_2$ line pointing directly out of the page and have been scaled to clearly show the dominant terms. Interpretation is as follows: (a) $N = 100$, here the double sum is dominated by the maximum at $n_1 = n_2 = 0$; (b) $N = 400$, here the double sum is dominated by a peak at large n_1 and $n_2 \approx 0$; (c) $N = 500$, here the double sum has two peaks that are similar in size at n_1 large, $n_2 \approx 0$ and $n_1 \approx 0$, n_2 large; (d) $N = 800$, here the double sum is dominated by the peak at n_2 large and $n_1 \approx 0$.

sites. As the number of defect sites is increased, however, this kind of behaviour may give over to metastability. It would be interesting to see if this is the case as this system could then shed some light on metastability, a phenomenon which is considered important in traffic models, as discussed in the following section.

3.5 Relation to traffic models

The ZRP has a long-known mapping to exclusion processes [3], where only a single particle can occupy a lattice site at any one time. The basic idea is to view the sites of the ZRP as occupied sites—or particles—of the exclusion process and the particles of the ZRP as vacancies of the exclusion process. Thus, if the ZRP has L sites and N particles, the exclusion process will have $L + N$ sites and L particles. In particular, an exclusion process with an interpretation in the area of traffic models known as the Bus Route Model (BRM) [15], has been studied. Here the mapping relates sites of the ZRP to buses in the BRM and particles in the ZRP to vacancies, or sections of road, which may contain passengers to be picked up.

Many other traffic models exist that are based on exclusion processes, see [30, 31] for some examples of these, along with discussion of real traffic and other related models. These exclusion process models have the road divided into a lattice of evenly sized road segments which can contain at most one vehicle. Furthermore, in the fundamental diagram, which relates traffic flux to global traffic density, overshoot-like behaviour is often seen in both the exclusion models and in measurements taken from real traffic, see Figure 3.13. Also this overshoot-like behaviour can occur around the transition point between freely flowing and jammed traffic phases.

A recent paper has further explored the relation between traffic models and the ZRP [33] by giving a direct mapping between the ZRP and a viable traffic model. Here the particles of the ZRP are related to vehicles of the traffic model and the sites of the ZRP are related to vacancies of the traffic model. Thus each occupied site of the ZRP represents a cluster of cars, i. e. a number of cars with no gaps between them. With this mapping the flux of cars can be related straightforwardly to the current of particles in the ZRP. Hence particle current data from the ZRP can be used to generate the fundamental diagram of a corresponding traffic model. An example of this is shown in Figure 3.14; similarities can be drawn between this and with a typical fundamental diagram taken from a traffic model, as shown in Figure 3.13. Thus the behaviours of the models can be connected, even though this particular ZRP traffic model is not based on exclusion principles.

In the ZRP, the usual hop rate from a site is one that decays with increasing

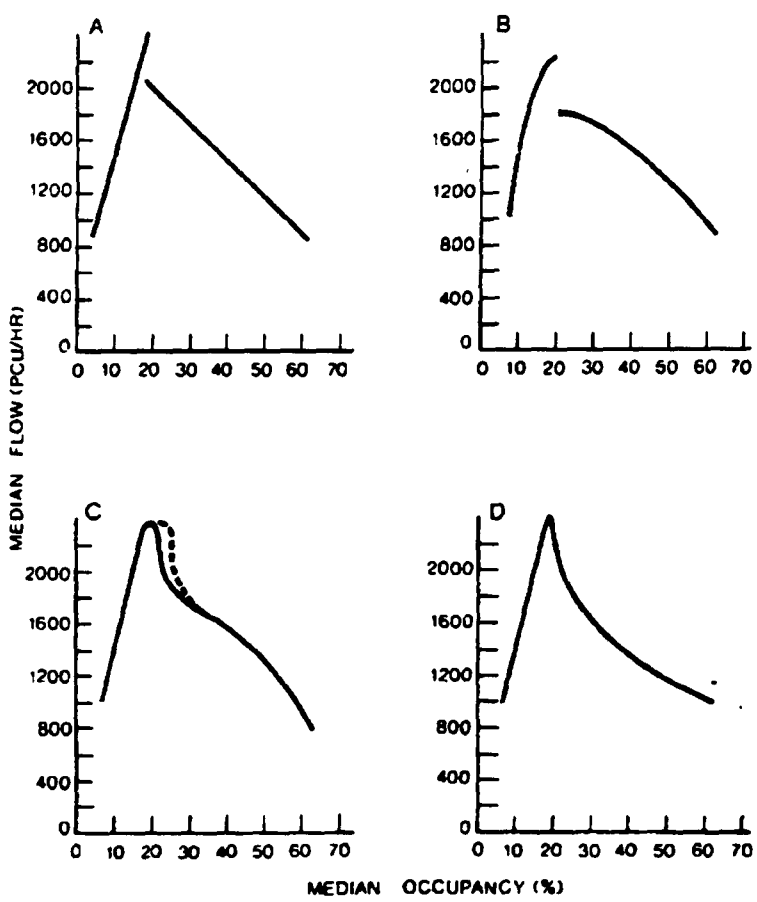


Figure 3.13: Several possible schematic forms for the current-density relation that are consistent with empirical observations (taken from [32]). It should be noted that several of these diagrams contain features that could be interpreted as an overshoot.

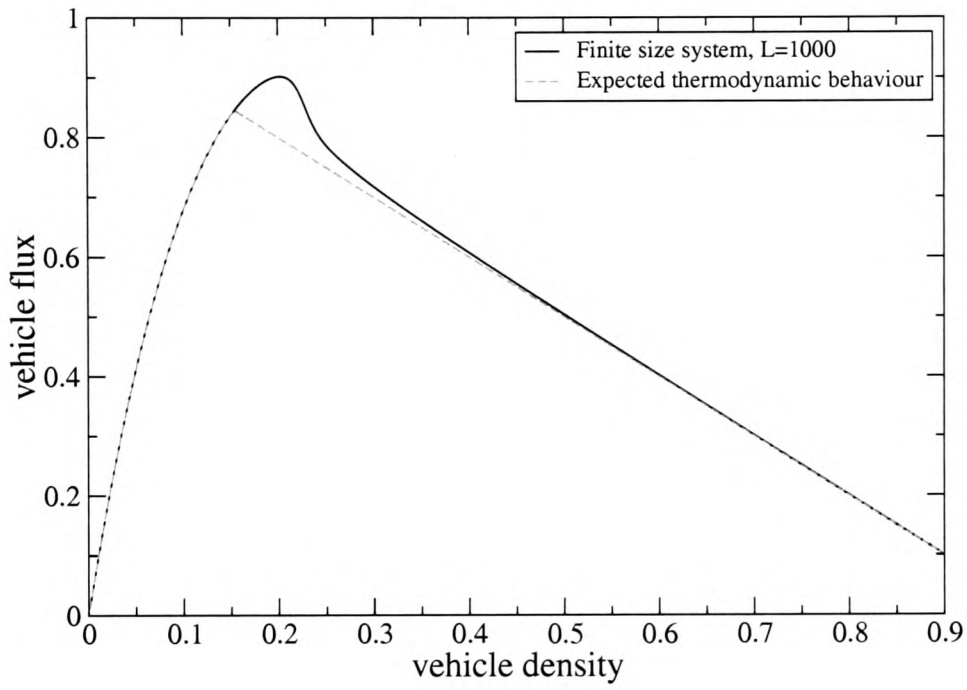


Figure 3.14: Fundamental traffic diagram, i. e. vehicle flux versus vehicle density for a homogeneous ZRP with hop rate $u(1) = 10, u(n) = 1 + 6/n$ for $n > 1$ (solid black line). Data shown are calculated from an exact numerical evaluation of the partition function for a system of size $L = 1000$. Also shown is the expected thermodynamic behaviour (dashed grey line). Comparing these data with the schematic fundamental diagrams from real traffic data shown in Figure 3.13 indicates that overshoot-like behaviour may be seen in real traffic.

particle number. Thus vehicles take longer to move away from clusters of vehicles and this principle is motivated as follows. Several traffic models use what are termed 'slow-to-start' or s2s rules, see for example [34, 35]. These rules model the fact that vehicles that have been stationary for longer times often take longer to start moving and may begin by moving slowly if expecting to stop again. Thus the decaying hop rates that give condensation in the ZRP have an interpretation as slow-to-start rules in the corresponding traffic model: vehicles take more time to escape from larger clusters as they will have been stationary for longer.

An important observation of many traffic models is the presence of metastable states, often around the overshoot-like region. At vehicle densities just above the overshoot region, it has been observed that simulations of traffic systems started in a evenly-spread fluid phase will not jam during the duration of the simulation, whereas artificially nucleated jams at the same density will persist for the duration. The ZRP traffic model [33] is capable of reproducing this kind of behaviour. However, this behaviour was not seen in the SDS system. This could be due to the inhomogeneity of the SDS system, the defect site presents a definite advantage to nucleation on this site. Also it is not clear how a defect site of the ZRP would be interpreted in a corresponding traffic model. Thus it would be interesting to study the homogeneous ZRP with hop rates that give clear and strong overshoots to see if there is any link between the overshoot and metastability in these cases. If metastability is found it would also be interesting to study under what circumstances overshoot behaviour and metastability coincide and how this links with the results seen in the SDS system. This could then provide insight into metastability and the fundamental diagram of traffic models.

3.6 Summary

In this chapter a finite size effect manifesting itself as an overshoot in the current density diagram was studied in a ZRP with a single defect site. The overshoot was found to be initially a continuation of the fluid phase, giving way to a metastable condensate with the stability increasing as the density is increased. Interestingly analysis in the canonical and grand canonical ensembles gives differing predictions on when the overshoot should be present and what form it takes. Analysis done in the canonical ensemble agreed well with results from simulations done in the canonical ensemble, which is the natural ensemble in which to simulate the ZRP, while a grand canonical analysis agreed less well. Thus, the canonical ensemble is the more accurate when applied to a finite ZRP such as that considered in this chapter. This should not

necessarily be taken to indicate a deficiency in the grand canonical ensemble, it is just that the two ensembles differ for systems of finite size. This is something that is often seen to be the case.

A ZRP with two defect sites was also studied and it was found that the two slow sites could produce interesting behaviour where a condensate can initially begin to form on one of the sites before switching to the other at higher density.

The general relation between ZRP models and traffic models was also discussed and overshoot-like behaviour seen in real traffic data was compared to that seen in the ZRP.

Some of the results of this chapter have been published in [36].

Chapter 4

Networks

4.1 Introduction

There are a great many systems in nature that are composed of distinct entities that interact with one another in some way. In many cases mapping the structure of these interactions is a useful analysis of the system. The maps of these interactions are known as networks which are comprised of nodes to represent the entities and links to represent the interactions. Examples span many branches of science and include: Social networks, for example the network of friendships [37] in a class of school-children where the nodes are children and a link is a friendship between two children. Biological networks, for example metabolic interaction networks where metabolites are nodes and links connect metabolites that are produced from one another in reactions in an organism [38]. Information networks, for example scientific collaboration networks [39, 40, 41, 42, 43] where scientists are nodes and a link connects scientists that have collaborated in a publication. Technological networks, for example the World Wide Web (WWW) where web pages are nodes and hyperlinks are the links [44, 45, 46, 47, 48, 49]. There are even networks from physics problems, for example the network of a potential energy landscape [50] where the nodes are local minima in the free energy and the links are saddle points which connect these minima. There are of course many more such examples, details of some of which and more involved discussions of many of the concepts introduced in this chapter can be found in the multitude of books and review papers on this subject [6, 7, 8, 9].

There are many ways to map such interacting systems: the nodes and links can carry a lot of information. The most common and most basic way is to consider all nodes as being the same, i. e. carry no information other than which entity of the system they represent, and all links to be binary, i. e. either present or absent. In

more mathematical terminology such a network is a graph [51] with a set of vertices being linked by a set of edges. A simple example of such a network or graph is shown in Figure 4.1. The terms network, nodes and links are more common in the physics literature (although sites and bonds are also used) whereas in the mathematics literature usually only the terms graph, vertices and edges are used.

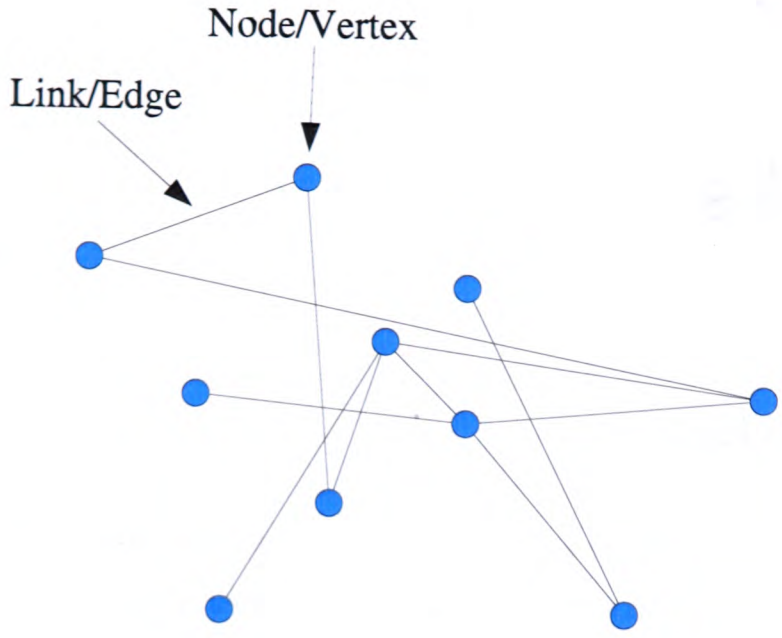


Figure 4.1: A simple graphical representation of a network. The circles represent nodes or vertices and the lines represent the links or edges that join them.

4.2 Complex networks

Recently, techniques from physics, and in particular statistical mechanics, are being applied to study complex networks, see [6, 7, 8, 9] for reviews. Network research has been around for a long time, but it is only recently that computer power and storage have become sufficiently advanced to analyse networks of a size which give good statistics and where the ideas of statistical physics may come into play. The advances in computer technology have also led to large new networks being created, some examples being the Internet and the World Wide Web. Where previously only small data sets had been analysed and network theory was just that, a theory with little concrete empirical backing, now statistical properties of large real data sets were possible. With large data sets, ways of characterising a network by statistical properties became more necessary—previously it had been possible to simply glean many properties of a network by eye from a diagrammatic representation. Some of the key properties of interest are:

- **Degree**—The degree, k_i, of a node, i, is the number of nodes that are linked to it, i. e. the number of nearest neighbours of node i. The distribution of degree amongst the nodes in a network or ensemble of networks, i.e. the degree distribution, $\mathrm{P}(k)$, is an important characteristic of a network.

- **Path length**—The path length between two nodes is the smallest number of links that must be traversed to get from one to another. The average path length, ℓ is the path length averaged over all pairs of nodes in a network.

- **Clustering**—The tendency of a network to highly connected groups of nodes, or clusters, can be measured with the clustering coefficient, C. The clustering coefficient is defined as the average over clustering coefficient of an individual node i:

$$C_i = \frac{E_i}{k_i(k_i-1)/2}\,, \tag{4.1}$$

  where E_i is the number of connections that exist between the nearest neighbours of node i and $k_i(k_i-1)/2$ is the number of possible links between the nearest neighbours. An equivalent and more intuitive definition of the clustering coefficient is the probability that if node i is connected to nodes j and k then nodes j and k are also connected to each other.

Various other properties have been proposed that characterise a network, for example there are: various 'betweenness' quantities that characterise how important certain nodes are in passing information across a network [52, 53]; degree correlations, which measure the likelihood that a node with degree k connects to a node with degree m [54, 55, 56, 57]; connected component sizes, which measure the number of nodes in a group in which all nodes can be reached from all others in some number of steps [58]. There are also countless other specialised properties that are not listed here.

With the scope of networks study being so vast, there are many networks with widely varying values of the above properties, however, there are a surprising number whose behaviour falls into narrow categories for those properties that are highlighted above.

- **Scale-free networks**—Part of the surge of recent interest in the study of networks has been due to the discovery that many networks have a broadly-tailed degree distribution, i. e. one that decays more slowly than an exponential. Many networks have shown a scale-free distribution, by which it is meant that the degree distribution has a power-law tail. The term scale free comes from the fact that in a power-law there is an absence of any characteristic scale. Scale-free degree distributions are also thought to be strong indicators of other properties

in a network. Scale-free degree distributions have been measured in the WWW for example, as reported in [44, 49, 58].

- **Small-world effect**—Many networks have been found to have a relatively small average path length when compared to the number of nodes or links in the network. In fact it was observed that many networks have an average path length that goes like the logarithm of the number of nodes in the network, $\ell \sim \ln N$. Networks that have an average path length that decays at least as slowly as this with the number of nodes are said to have the small-world property. This kind of effect was first popularised by Milgram [59] who asked people to get a letter to someone they did not know by sending it to a first-name basis acquaintance, asking them to pass it on in a similar fashion. It was found that, on average, only six such steps were required to reach the target — this has come to be known as 'six degrees of separation'. The small world property is important for electronic communication networks as a small path length generally coincides with a short message delivery time or overhead. The small world effect has also been discussed for the SCN in [39].

- **High clustering**—In social networks if person A knows persons B and C, then it is highly likely that persons B and C also know one another. Thus social networks will have a high clustering coefficient. Furthermore it has been observed that as network size increases, with a constant average degree, the clustering seems to tend to a constant limit. While not as high as in social networks, the clustering of many biological, technological and information networks is also large and the coefficient again seems to be tending to a constant value as network size increases. A 'high' clustering coefficient is generally taken to be one that is higher than a random graph—these will be discussed shortly in Section 4.4.1. Clustering is also often referred to as transitivity in the literature. High clustering has been measured in many real networks, as reported in [60, 61].

The three properties highlighted above are probably the most commonly studied. In fact many studies look primarily at the degree distribution and it is that with which this body of work is mostly preoccupied.

4.3 Generalisations

The most basic networks only map the connections between labelled, but identical, entities with simple links. More generally a network could be composed of different

entities that are connected by different kinds of links. The following generalisations of basic networks go some way in accounting for this:

- **Directed networks**—In many real networks the links have a concept of direction, for example in the WWW a hyperlink may link may link page A to page B, but page B may have no hyperlink back to A; it would not be possible to get from B to A without knowledge of the URL of A. Thus networks in which directed links are commonly studied. Generalisation of many of the properties highlighted above is straightforward. Directed networks will be discussed in more detail in Chapter 6.

- **Weighted networks**—In addition to the concept of direction, links may also differ in importance. For example, in social networks a link may represent a casual acquaintance or a lifelong friend. Clearly the latter will be more relevant than the former in many properties of the network. As such, links can be assigned a weight which represents the importance of the link. Weighted networks will also be discussed in more detail in Chapter 6.

- **Bipartite networks**—Networks may also have different types of node. For example, the scientific collaboration network (SCN) [39] can be described by a network consisting of authors and papers, with authors only linked to papers which they have (co-)authored. Collaboration information can then be gleaned from the unipartite projection of this, where authors are connected if they each have a link to the same paper. The bipartite picture is often favoured as taking the unipartite projection disposes of some of the information. Generalisations to multipartite networks are also possible.

- **Spatial networks**—Thus far no mention has been made of how the position of a node in real space can affect things. In Figure 4.1 the position of the nodes is really for convenience and carries no meaning the basic representation of the network would be the same if the nodes were moved around without changing which nodes are linked. For abstract networks like the WWW the position of a web page does not really mean anything, any web page can link to any other as long as the URL is known. However, in the Internet the position of a node has meaning as major nodes will tend to be located where people, businesses and/or institutions are [62]. Another example of a network in which spatial constraints is important is the world-wide airport network (WAN), where nodes represent airports and links direct flights. Clearly for these networks there is a limit to how far a direct commercial flight can be and also the location of airports is

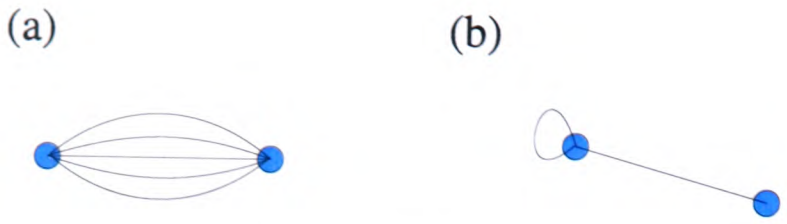


Figure 4.2: Examples of (a) repeated connections or melons and (b) self connections or tadpoles.

often restricted by geo-political considerations [63]. Fortunately, information about the position of a node can easily be encoded in the labelling of the nodes.

Within networks and network models there is also the issue of repeated and self connections, i. e. multiple links between the same pair of nodes and links which connect a node to itself. Although not strictly a generalisation, these concepts become important when modelling networks and in some real networks themselves. In modelling networks it is often easier to work with models that allow for multiple and self connections, also known as melons and tadpoles—see Figure 4.2. Also in the WWW one may have several repetitions of the same hyperlink in a web page and also hyperlinks that point to the same page they are on, for example a home link is quite often present on the root page as well as the sub-pages. The effect of the presence or absence of repeated and self connections will be discussed in more detail in Chapter 5.

4.4 Network models

Networks often form important backbones in complex systems and their study can be the key to understanding a system. As discussed above, networks can be characterised by various quantities and many networks show strikingly similar values for some of these characteristics. It is desirable to know what, if any, organising principles cause networks to form with these values. As such many models of networks based on simple rules have been proposed and studied and have given some insight into how and why networks come together with the observed properties. Some of the seminal models of networks are discussed in the following sections.

4.4.1 Erdös-Rényi random graphs

One of the earliest models of complex networks is the random graph, dating from the 1950s. Although studied independently by other authors [64], and in fact published earlier, random graphs are most commonly associated with Erdös and Rényi [65] who

performed extensive and rigorous studies. Random graphs have the advantage of being very simple and are also intuitive to the statistical physicist.

In the original definition of the Erdös-Rényi random graph model (ER-RG) a fixed number, N, of nodes were connected by n links randomly chosen from the $N(N-1)/2$ possible links. An equivalent definition in the thermodynamic limit, and one that is more in keeping with the concepts of statistical physics, is to connect every possible pair in a set of N nodes with probability p, with p chosen such that $n = Np$. A very simple network constructed by such a procedure is shown in Figure 4.3.

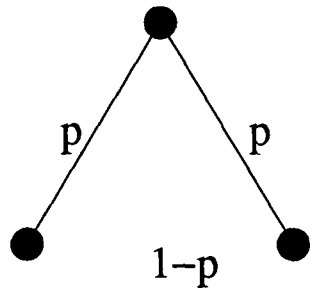


Figure 4.3: A very simple Erdös-Rényi Random Graph. Two nodes are linked with probability p and not linked with probability $1 - p$.

As can be seen the model is quite simple and a great many properties of it can be calculated exactly, in depth of random graphs and what calculations are possible can be found in many of text books on the subject, for example see [66]. The probability that a node has degree k is simply the probability that k of the $N - 1$ possible links from it are present

$$\mathrm{P}(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} , \tag{4.2}$$

giving the ER-RG a binominal degree distribution. This approximates closely a Poisson distribution and this is often quoted as the degree distribution of a random graph—of course it does become exact in the limit of large, sparse networks. The average path length is less easily calculated, but it can be rigorously shown [66, 67] that the average path length behaves as

$$\ell = \frac{\ln N}{\ln \langle k \rangle} , \tag{4.3}$$

where $\langle k \rangle$ is the average degree in the network which is equal to Np. Thus, as with many real networks, random graphs exhibit the small-world property. The clustering coefficient is easily calculated: if person A knows persons B and C, then the probability that B knows C is simply p and is in fact independent of A knowing either B or C, thus

$$C = p = \frac{\langle k \rangle}{N} . \tag{4.4}$$

Hence, when considering random graphs with fixed, finite average degree, the clustering coefficient is small and will decrease to zero in the limit of a large number of nodes. When it is said that the clustering coefficient of a network is 'high', it is often meant that the clustering coefficient is larger than that of the random graph with the same number of nodes and average degree.

Such random graphs have long been used to model real networks, mainly due to their simplicity and a lack of real networks data for comparison. Their main success is that they possess the small-world property as is in keeping with real networks. However, the degree distribution differs drastically with that observed for many real networks in that it does not possess a broad tail. Also the clustering coefficient is generally much lower than those observed for real networks and tends to zero in the limit of large network size for fixed degree in contrast to the finite constant limit apparently seen for real networks.

Random graphs with arbitrary degree distribution

A step that can be taken to rectify the failure of random graphs to accurately model degree distributions of real networks, is to constrain the degree distribution to have the desired form from the outset. Essentially graphs that are maximally random subject to the constraint that the degree distribution is fixed are studied. Newman, Strogatz and Watts have studied the calculation of various properties of such networks using a generating function formalism in [68]. More rigorous calculations of some properties have been done by Dorogovtsev, Mendes and Samukhin in [69]. Further rigorous calculations have been done in the configuration model, in particular in the work of Molloy and Reed [70, 71]. In the configuration model a particular degree sequence is drawn from the distribution; in the limit of large system size this should be representative of the system. With a given degree sequence each node has a number of stubs, i. e. ends of links, and these are then connected together at random to create full links, under the constraint that no multiple or self links are created. This generates every possible graph for a given degree sequence with equal probability. A similar model has been proposed by Chung and Lu [72, 73], in which instead of choosing a specific degree sequence, an expected degree sequence is specified. Here each node is assigned an expected degree and nodes are then linked with a probability that depends on the expected degree of each node. This has the disadvantage that the degree distribution generated is not always exactly the same as the expected degree distribution, however many calculations are easier within this framework and is some sense it is a more natural generalisation of the random graph.

Such generalised random graph models are useful as null models of scale-free networks. They reproduce the small-world property and have desired degree sequences, but still fail on reproducing the large levels of clustering seen in many real networks, although more recently variations which fix the clustering as well as the degree distribution have been studied [74]. Also it is often seen as a disadvantage that the degree distribution has to be specified, one of the aims of network models is to gain insight into how such degree distributions arise and in this respect constraining the degree distribution at the outset is not satisfactory.

4.4.2 Watts-Strogatz small-world networks

In an attempt to improve on random graphs a model was proposed that aimed to keep the mechanism in random graphs that generated the short path length whilst having a separate mechanism to generate clustering. To do this Watts and Strogatz formulated a model that interpolated between a regular lattice and a random graph [75]. The appropriate regular lattice having high clustering and the random graph having the small-world property.

The model is defined as follows:

1. Start with a ring of N nodes, each connected to their 2κ nearest neighbours (κ on each side). See Figure 4.4 (a) for an example with $\kappa = 2$.

2. Rewire each link present in the initial ring with probability p, connecting a randomly chosen end to another randomly chosen node, see Figure 4.4 (b).

The driving idea behind this model is that the ring lattice with $\kappa > 2$ is highly clustered and that the random rewirings introduce long range connections that reduce the average path length. When $p = 0$ the small-world network (SWN) is a regular lattice; it has high clustering $C = 3(\kappa-1)/2(2\kappa-1)$ for $\kappa \geq 2$ [76], but a large average path length $\ell = N/4\kappa$. On the other hand, when $p = 1$ all the links are randomly rewired and the result is something very similar to a random graph and so expected to have low clustering and a path length with the small-world property $\ell \sim \ln N$. Fortuitously it turns out that only a relatively small number of long-range links are needed to ensure the small-world nature of the resulting network. In fact this number is small enough that there is a range of values of p for which the networks is both highly clustered and small-world in nature.

The relevant properties of the Watts-Strogatz Small World Network (WS-SWN) model have been studied in detail in [76]. They find that the degree distribution

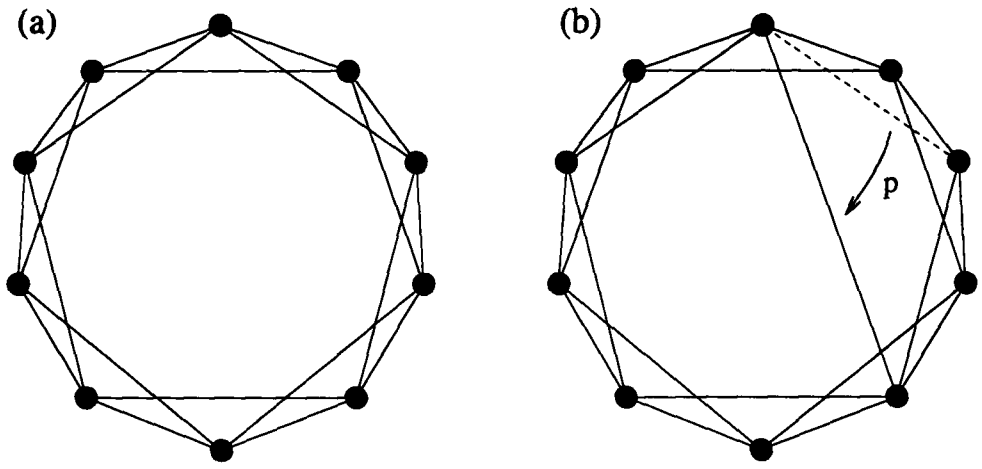


Figure 4.4: Graphical representation of a single rewiring event for the Watts-Strogatz Small-World Network model. (a) the initial configuration of the network, a ring of nodes with each connected to its $\kappa = 2$ nearest neighbours on each side. (b) a single link rewiring event; the link denoted by a dotted line has been removed and rewired at one end with with probability p.

behaves as

$$P_p(k) = \sum_{n=0}^{\min(k-\kappa,\kappa)} \binom{\kappa}{n} (1-p)^n p^{\kappa-n} \frac{(p\kappa)^{k-n-\kappa}}{(k-n-\kappa)!} \exp(-p\kappa) \,, \tag{4.5}$$

for $k \geq \kappa$ and zero otherwise. This is a little complicated, but a similar model is to start with the ring and add links with probability p and in this case it is obvious that one gets a binomial distribution. Thus it is fair to compare the degree distribution of an SWN to that of a random graph. The average path length was found to follow a scaling relation

$$\ell(N,p) \sim N^* F_\kappa \left(\frac{N}{N^*} \right) \,, \tag{4.6}$$

where F_κ depends only on κ and satisfies

$$F_\kappa(x) \sim \begin{cases} x & \text{if } x \ll 1 \\ \ln x & \text{if } x \gg 1 \end{cases} \,. \tag{4.7}$$

Thus if the network is well above some p-dependent size N^*, or equivalently is well above some N-dependent p value, then it will have the small world property $\ell \sim \ln N$. The clustering coefficient is found to be

$$C(p) = \frac{3(\kappa-1)}{2(2\kappa-1)} (1-p)^3 \,, \tag{4.8}$$

with corrections of order $1/N$. As stated above, for a fixed network size, the value of p for which the WS-SWN takes on the small-world property is generally lower than that for which the clustering becomes appreciably different from the underlying regular

lattice. Thus the WS-SWN succeeds in terms of both clustering and the small-world property, but like the ER-RG model it fails to give a power-law degree distribution.

4.4.3 The Barabási-Albert model

The models discussed above are all examples of what are often termed static models: they have a fixed number of nodes and while they may have some sequential construction procedure, this is something that is done once and then the network is fixed. These static models were successful in describing some aspects of real models, but a realistic degree distribution could only be obtained by imposing it on the system. Ideally a model should get a realistic degree distribution from a set of simple rules, this model should then give insight into the mechanisms via which broadly-tailed degree distributions arise. With this in mind Barabási and Albert proposed a simple *growing* model [77]—now known as the Barabási-Albert (BA) model—which gives a power-law degree distribution from simple rules.

The model is defined as follows: Starting from an initial seed network of m_0 nodes (for the asymptotic behaviour the connections in the seed network are not important) the network evolves via a process with two main elements:

1. **Growth**—At each discrete, evenly-spaced time step a new node is added to the network with $m \leq m_0$ links attached.

2. **Preferential attachment**—These links are then attached to existing nodes of the network with a probability

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}\,, \tag{4.9}$$

where k_i is the degree of the node to be attached to. In fact this is *linear* preferential attachment as the probability of attachment is linear in the degree of the node.

These rules were motivated by concepts from several real systems. Many networks are clearly growing entities, for example the World Wide Web and the Internet are both currently growing in size and show little sign of stopping. Also the idea of preferential attachment has long been around in the sociology literature [78] and has physical basis in several network models, for example in the science citation network papers that are already well cited are likely to attract further citations, both because they are more likely to be found when following citations from other papers and in that having many citations, they are likely to be of high quality. The idea of preferential attachment is also seen in models of wealth distribution and has generally gone under

several monikers including "rich get richer", "cumulative advantage" and "popularity is attractive".

As stated previously the BA model yields a power-law degree distribution. An approximate solution due to Barabási and Albert [77, 79] shows the degree distribution to be

$$\mathrm{P}(k) \sim k^{-\gamma} , \tag{4.10}$$

with $\gamma = 3$. This exponent is reasonable as many real networks have had degree distribution exponents measured as being in the range $2 - 3$. The degree distribution can also be solved exactly, in the thermodynamic limit, via a rate equation method [80] and a similar master equation approach [81] yielding the same asymptotics. The degree distribution of a similar, more rigorously defined model has been calculated in [82]. The average path length and clustering coefficient have not yet been calculated exactly, but extensive numerical simulations have been performed. It was found that the average path length data was fitted well by the form $\ell = A \ln(N-B)+C$ with A, B and C constants [6]. This implies that the small world property $\ell \sim \ln N$ is satisfied by the BA model, however some analytical work has predicted that there may be a double logarithmic correction to this $\ell \sim \ln N / \ln \ln N$. [67, 83]. The clustering coefficient has also been studied numerically, but as yet no analytical predictions have been made. It was found [6] that the clustering coefficient of the BA model is generally higher than that of a random graph model with the same average degree for different system sizes, but that the clustering was decreasing with N as $N^{-0.75}$, which, while a slower decay than for a random graph, is still different from the absence of decay seen in many real networks. Thus the BA model yields reasonably realistic degree distribution behaviour and the small-world property, but it fails on clustering, albeit not as badly as the ER-RG model.

One of the major differences between this model and the previous two is that the this model is growing and so the number of nodes and links increases with time. This gives this model an inherently nonequilibrium nature. Thus it may be expected that techniques from nonequilibrium statistical mechanics may prove useful in the analysis of this and other network models.

4.4.4 Generalisations of the BA model

The BA model is one of the simplest models that can reproduce power-law degree distributions from intuitively reasonable simple rules. It was seen in the previous section that, in its basic form, the BA model generates a scale-free network, i. e. one with a power-law degree distribution, with exponent $\gamma = 3$. Of course, real scale-free

networks do not all have this exact exponent, many have been measured in the range 2-3 and several lie outwith this range. Many generalisations of the basic BA model exist which improve this situation. As some of the concepts of the BA model and its generalisations will be central in some of the following sections, some of the simplest generalisations are discussed in detail.

Asymptotically linear preferential attachment

The original BA model had linear preferential attachment for all degree. Since only the tail of the degree distribution needs to be a power law for it realistically represent a real network, a simple and logical generalisation is to make the preferential attachment only an asymptotically linear function of the degree [80, 84, 47], i. e.

$$\Pi(k_i) \sim k_i \ . \tag{4.11}$$

It is found that the non-linear behaviour for small degree can significantly affect the degree distribution. For example, even taking $\Pi(1) = a_\infty \mu$ and $\Pi(k > 1) = a_\infty k$ (where a_∞ is a normalisation constant) [80, 84] results in a scale-free degree distribution with an exponent that can be tuned in the range $(2, \infty)$ simply by varying the constant μ.

Additional attractiveness

Another related way to generalise the BA model is to give each node a uniform additional attractiveness [81, 84, 85], i. e.

$$\Pi(k_i) = \frac{(A + k_i)}{\sum_j (A + k_j)} \ , \tag{4.12}$$

where A is a constant. Again only a slight change to the pure linear preferential attachment rule can profoundly alter the degree distribution. Here it is found that when varying the additional attractiveness, A, and/or the number of links each node has when it is first introduced, m, the degree distribution retains a power-law form with an exponent that, as in the previous case, can be tuned in the range $(2, \infty)$. The effect of additional attractiveness has also been studied for directed networks in [86, 84], where both the in- and out-degrees can be similarly tuned.

The above generalisations by no means exhaust the catalogue of generalisations of the BA model. Some of the other—sometimes seemingly innocuous—generalisations can have an even greater effect on the basic model than to tune the exponent of the degree distribution.

Non-linear preferential attachment

As stated above, one of the principal ideas behind the BA model is that of preferential attachment. In the original model only linear preferential attachment was considered. It was seen above that even a slight change to the non-asymptotic properties of the attachment probability could have a strong effect. The obvious next step is to consider preferential attachment that is non-linear for all degree,

$$\Pi(k_i) = \frac{k_i^\alpha}{\sum_j k_j^\alpha} , \tag{4.13}$$

with $\alpha \geq 0$ was first studied in [80] for $m = 1$ and subsequently in [84] and in [47] for directed networks. It was found that: for $\alpha < 1$ the degree distribution has a stretched-exponential tail; for $\alpha = 1$ the BA model is recovered along with the power-law tail in the degree distribution; for $1 < \alpha \leq 2$ a single node captures a finite fraction of all the links in the system, even in the infinite limit—a phenomenon known as condensation or gelation; finally for $\alpha > 2$ a single site has a non-zero probability of being connected to all other sites. More recently the case of $m > 1$ with non-linear preferential attachment has been studied in [87], which also looked at the clustering coefficient and average path length. Here, broadly the same behaviour for the degree distribution was found, with the exception that in the gel phase there were m super-connected nodes. It is interesting to note that only linear preferential attachment yields a power-law degree distribution. A discussion of the validity of this choice in terms of real networks is given in Section 4.5.

Inherent attractiveness, fitness

Thus far only homogeneous systems have been discussed in that the functional form of the attraction is the same for all sites, the idea of preferential attachment is that the richest node in terms of connectivity is the one that attracts the links, but it seems feasible that some nodes may simply be more able to acquire links irrespective of their degree. To this end networks with some intrinsic fitness of nodes have been studied, with so-called fit-get-richer rules. Models with multiplicative fitness, i. e. a new node attaches to an old node, i, with probability

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j} , \tag{4.14}$$

where η_i gives the fitness of node i, were first studied in [88, 89]. The fitness for each newly added node is drawn from some distribution $\rho(\eta)$. In [88] the network, in its thermodynamic limit, was mapped onto a Bose gas. In that work the fitness of a node was related to an energy, hence relating nodes and energy levels and the number

of connections a node has was related to the occupation of an energy level. Upon introducing a temperature parameter in the relation between energy and fitness it can be shown that there exists a temperature dependent transition between a fit-get-rich phase, where the higher fitness nodes get the most links and the degree distribution has a power-law tail; and a Bose-Einstein condensate phase, where the node with the highest fitness captures a finite fraction of all the links in the system. Subsequently multiplicative-fitness-driven behaviour has been studied in [85, 90]. Additive fitness (also sometimes known as additional attractiveness), i. e. a connection probability given by

$$\Pi_i = \frac{k_i + \eta_i}{\sum_j (k_j + \eta_j)} , \tag{4.15}$$

has been studied in [85, 91]. It was found that this can yield a power-law degree distribution with an exponent that depends on $\langle \eta \rangle$. A mixture of multiplicative and additive fitnesses was looked at explicitly in [91], it was seen that this can yield a degree distribution that follows a power-law with an inverse logarithmic correction, as was also seen for multiplicative fitness alone in [89]. The case of fitness without preferential attachment was studied in [92, 93]. Here, in a static construction, vertices are assigned a fitness from some distribution and then linked with some probability that depends on the fitness at both ends of the possible link. It was found that scale-free networks can be generated for any distribution of fitnesses, provided the correct linking probability is chosen. A similar model was studied in [94] and such static fitness models have been found to belong to the more general class of networks with hidden variables [95].

Non growth driven rewiring events

Another simple generalisation of the BA model is to allow the linking of the system to change not just by the linking of new nodes to the system, but by other mechanisms in parallel with this. For example, in [96] at each time step one of three processes occurs: Either with probability p, m new links are distributed among the existing network—one end chosen randomly, the other with linear preferential attachment. Or with probability q, m links are rewired—a link from a randomly chosen node is moved to another with linear preferential attachment. Or with probability $1 - p - q$ a new node is added to the system with m links and these are attached to existing nodes with linear preferential attachment. Under these rules, depending on p, q and m the system can find itself with an exponential degree distribution or a scale-free distribution with exponent in the range $2 - \infty$. Many more variants of growing models with the link structure not solely created by the addition of new nodes to the system have been

considered in [81, 97, 48, 85, 86, 98, 99, 91, 42, 47, 90, 100], for example.

Other generalisations

Some examples of other generalisations include the following:

- Ageing of nodes [101, 85], where nodes have a multiplicative fitness that decreases in time, leading to a scale-free degree distribution with an exponent in the range $(2, \infty)$.

- Accelerated growth [101, 85, 102], where at each time step a node is added and connected along with a number of extra links with the number of extra links added an increasing function of time. This models the fact that the number of links in the Internet and WWW are increasing at a faster rate than the number of nodes. This can lead to scale-free degree distributions with an exponent that is less than 2.

- Death of nodes [85, 103], where at each time step a randomly chosen existing node is deleted with some probability. It turns out that asymptotically this makes no difference to the degree distribution. This is starkly different to the case with random deletion of links which was found to alter the degree distribution exponent [97]. It was also found that preferential deletion of highly connected nodes removes the scale-free nature of the degree distribution. Many studies have been made of random failure of nodes and intentional attack against the most connected nodes in the BA model and others, and in real networks, for some examples see [104, 105, 106, 107, 108, 109] where it was generally found that scale-free networks are robust against random failure, but vulnerable to intentional attack.

- Active and inactive nodes [110, 111], where m nodes of a network are active and each receives one of m links from a newly added node. The newly added node then becomes active and one of the active nodes is deactivated with a probability that decreases with degree. This leads to highly clustered networks with the small-world property and scale-free degree distributions. However, this mechanism is not intuitively realistic and this has not been generally accepted as one of the underlying principles behind the behaviour seen in real networks.

- Link copying, walking and redirection rules [45, 112, 46, 84, 113, 114, 115, 116], where, for example, nodes will link to other nodes, often chosen at random, but then redirect to point to a nearest neighbour of this node or copy a number

of links from this node or undergoes such a process recursively. Such rules have been found capable of producing scale-free networks with a wide range of exponents and sometimes with increased clustering.

4.5 Validity of preferential attachment

The mechanism of preferential attachment is one of the founding concepts behind the BA model and its derivatives, and in the preceding section it has been seen that for preferential attachment alone, *linear* preferential attachment is required to generate scale-free networks. It has also been speculated that the scale-free structure of a network implies that it has grown through some linear preferential attachment mechanism [117]. The concept of preferential attachment is also important in much of what follows. Natural questions which arise are: Is there some underlying mechanism that itself leads to linear preferential attachment and can preferential attachment be measured and verified as linear empirically? Some progress has been made in answering both of these questions. In the latter case empirical observations have been made of the time evolution of several real networks [118, 54, 55, 119, 42, 120] and a general method for reverse engineering of linking preferences from empirical measurement is discussed in [121]. It was found that several networks do indeed seem to evolve due to preferential attachment, but not always of the linear variety, usually being sub-linear in form. For those not growing through linear preferential attachment the degree distribution should not be a power law [80], however it has been speculated that another mechanism following linear preferential type rules could restore the scale-free behaviour, for example the creation of links between nodes at times after the entrance of each into the network [119, 42]. In the former case a few alternative rules that lead to a form of linear preferential attachment indirectly have been proposed [84, 113, 116] all of which seem reasonable and lend extra credence to the notion of linear preferential attachment. It has also been shown that a random walker on a general network will be found at a given node with a probability proportional to the degree of that node in the steady state[122].

4.6 Statistical mechanics of networks

The three models looked at in detail in section 4.4 were all successful at modelling some aspect of real networks and all contained ideas that hinted that the study of networks was in fact a field of statistical physics. The simple network models presented in the preceding sections tacitly involved ensembles of networks: a single network can

be created by a set of rules, but the calculated properties involve averaging over all networks created by these rules. Indeed the majority of network models are studied in this way. Recent work has attempted to elucidate the connection between the modelling of networks and statistical mechanics. For example, in the original definition of the ER-RG model where n links are chosen at random from the $N(N-1)/2$ possible between N nodes, each graph with n links is equiprobable. This can be mapped onto a classical statistical mechanics system, where each graph is a separate microstate and they all have the same energy and probability, in an analogue of the microcanonical ensemble.

There are several approaches when constructing a more formal statistical mechanics of networks. Each possible network configuration can be assigned an energy and the usual microcanonical, canonical and grand-canonical ensembles can be realised. The concept of energy need not even be introduced, a statistical weight can simply be assigned to each possible network configuration through some rule, and appropriate ensembles studied. A third option is to define some restructuring dynamics for the network and study the statistical properties of networks in any steady state that may be reached. This third option of defining a dynamics is a common starting point for nonequilibrium models and thus this approach allows the study of nonequilibrium models of networks, without being confined to nonequilibrium systems.

One of the earliest papers explicitly looking at statistical ensembles of networks [123] had little mention of energy. Instead the ensemble was defined through a partition function given by a minifield theory—a toy field theory in zero dimensions—and the weight of each network configuration was equated to the weight of the equivalent Feynman diagram. The couplings of the Feynman diagrams were chosen to give a scale-free network in which the allowed configurations were degenerate (i. e. with self and multiple links) connected tree graphs. Interestingly this model was found to map onto the balls-in-boxes model [124, 125]. It was found that, for certain parameter sets, scale-free degree distributions exist on a critical line between generic trees where all nodes have similar low degrees, and 'crumpled trees', where large star formations are seen, i. e. a few nodes have very many connections. The analysis was extended to non-degenerate graphs in [126].

Another work defined a network ensemble without going into field theory [127]; the ensemble was again defined by a partition function which was the sum over all network configurations with a fixed number of nodes and links, with each configuration weighted with a Boltzmann factor with a general Hamiltonian. The work in [123, 126] can be interpreted in a similar way, there the Hamiltonian was additive over nodes and only depended on the connectivity of the node—this lead to uncorrelated scale-free

networks. The work [127] looked at more complicated Hamiltonians that generated correlations between the degrees of the nodes, but failed to generate power-law degree distributions. The work in [123, 126] was extended to treat clustering in networks in [128, 129] and summarised in [130].

A similar ensemble approach was used in [131, 132] with both single vertex and many vertex energies to study topological phase transitions. For energies that depend on the size of the largest connected component a transition between disordered and giant component phases can be identified. Particular single-node energies depending on the connectivity show transitions between dispersed states to states with one or more star structures, i. e. many low degree nodes all connected to the same high degree hub nodes. At the transition point between the dispersed and star-like states scale-free degree distributions can be observed. A review of this approach and the relation of others to it has been given in [133]. Very recently a Hamiltonian that favours a node linking to other nodes with higher degree than its own was proposed in [134]. Numerical simulations of this system indicated that, when varying the temperature, the system moved between a phase with a core of highly connected hub nodes at low temperature and a phase with random connections at high temperature. It was also observed that the degree distribution had a power-law tail around the transition point, albeit sometimes with an extra high degree piece representing the highly connected hub nodes.

Network ensembles can also be defined by starting from the founding principles of statistical mechanics [135]. Instead of making a direct assignment of energy, one can simply constrain certain quantities of the ensemble to have average values equal to those measured in a real network, then assign probabilities to each possible network configuration such that the Gibbs entropy is maximised subject to the constraints. Such a formalism produces what are known as exponential random graphs. The probability distribution of the ensemble is an analogue of the Boltzmann distribution, but can also be thought of as the least biased probability distribution that gives average values equal to those that are believed to be true, i. e. those that have measured empirically.

Of course when such equilibrium statistical ensembles are studied, it is always possible to define some dynamics—for example the Metropolis algorithm [136]—that will move between configurations and create a dynamical system for which the time average is equal to the ensemble average. Several bodies of work put the emphasis on the dynamical picture rather than the energy picture—see, for example, [137]. This allows the study of evolving networks without growth; a concept central to the BA model and its variants. Such studies are important as, although growth dominates the connectiv-

ity changes in networks such as the Internet and World-Wide Web, in some networks it is believed that dynamic rewiring of links is the dominant factor determining the connectivity changes. Such networks include protein interaction networks—see, for example, [138]. Equilibrium network ensembles can also be defined by some rewiring dynamics; an example of this is discussed in detail in the following section. Defining non-growing network models through their dynamics also opens up the possibility of studying networks that are both non-growing and nonequilibrium.

4.6.1 Equilibrium rewiring networks

The practice of defining simple rewiring rules that lead to an equilibrium network theory was perhaps most clearly and comprehensively developed by Dorogovtsev, Mendes and Samukhin [139, 8]. They studied a system with N nodes and L links, with both these numbers conserved. Starting from some random initial configuration at each time step a randomly chosen end of a randomly chosen link would be rewired to another node preferentially with its degree, k, with some probability given by a preference function $f(k)$, see Figure 4.5. This rewiring procedure will repeat indefinitely

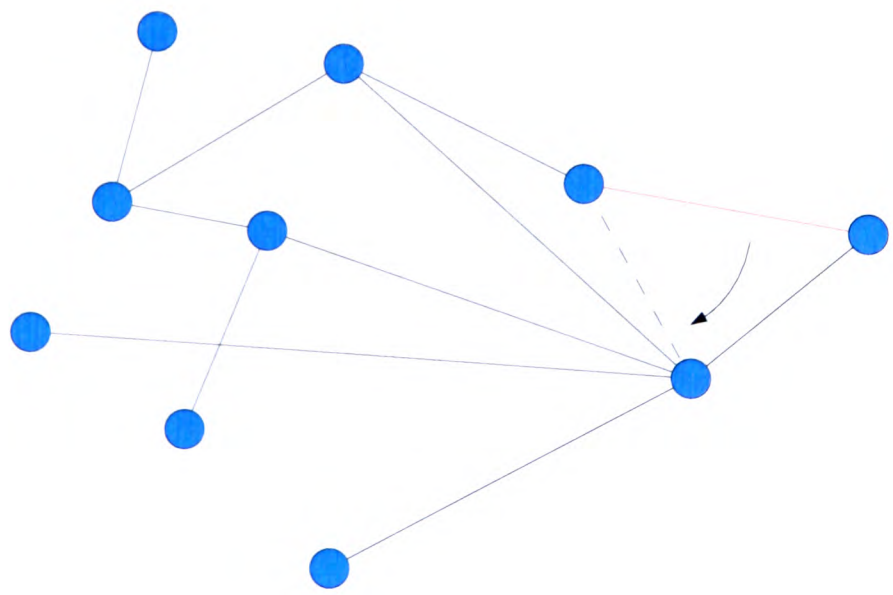


Figure 4.5: Rewiring move for the Dorogovtsev, Mendes and Samukhin equilibrium rewiring model. A link is selected at random (the selected link is highlighted) and rewired at a randomly chosen end to another node with a probability proportional to $f(k)$, which depends on the degree k of that node. The link after the rewiring move is shown by the dotted line.

and after a relaxation period the system will attain an equilibrium steady state. This can be proven [139] by noting that the dynamics define a Markov process and hence

conditions that guarantee a steady state are [140]: any state can be reached from any other in a number of steps; all processes are reversible and for any sequence of configurations that begin and end in the same configuration the probability of the sequence of moves between the states is the same as for the reversed sequence. These conditions can straightforwardly be shown to be true when the probability of a move depends only on the initial state and when the reverse of any move is possible, as are true for the dynamics given above. The probability of choosing a particular link and this rewiring to a node with a certain degree depend only on the initial state and as a rewiring can happen to any node all moves are reversible.

Thus this model combines the ideas of a equilibrium statistical physics in its definition of an ergodic dynamics to generate an ensemble and the idea of preferential attachment from the BA model, which was one of the first successes in obtaining a power-law degree distribution. The analysis of this model is discussed in some detail in the following as it has a strong connection with the ZRP to be discussed in Section 4.7.

The degree distribution is most easily obtained via a master equation approach [8]

$$\begin{aligned} N\frac{\partial \mathrm{P}(k,t)}{\partial t} = & \frac{1}{\langle f(k)\rangle}\left[-f(k)\mathrm{P}(k,t) + f(k-1)\mathrm{P}(k-1,t)\right] \\ & + \frac{1}{\bar{k}}\left[-k\mathrm{P}(k,t) + (k+1)\mathrm{P}(k+1,t)\right] , \end{aligned} \tag{4.16}$$

where $\mathrm{P}(k,t)$ is the degree distribution of the network at a given time t, $\bar{k} = 2L/N$ is the average degree of a node and $\langle f(k)\rangle$ is the ensemble average of the preference function $f(k)$. The first and third terms on the RHS of (4.16) represent the probability current out of a state with a node of degree k due to a rewiring event to and from a node with degree k respectively; the second and fourth terms represent the probability current into a state with a node of degree k due to rewiring to a node with degree $k-1$ and away from a node with degree $k+1$, respectively. The above equation is strictly only valid for $k \geq 1$. For $k = 0$ the following boundary equation applies

$$N\frac{\partial \mathrm{P}(0,t)}{\partial t} = -\frac{f(0)}{\langle f(k)\rangle}\mathrm{P}(0,t) + \frac{1}{\bar{k}}\mathrm{P}(1,t) . \tag{4.17}$$

The first term on the RHS of the above equation represents the probability current out of a state with a node of degree 0 due to a rewiring event to a node of degree 0 (note that there can be no rewiring event from a node of degree 0). The second term represents the probability current into a state with degree 0 due to a rewiring event from a node with degree 1. In the steady state $\mathrm{P}(k,t) \to \mathrm{P}(k)$, and the LHS's of the

two previous equations go to zero. Thus the following steady state condition is found

$$\begin{aligned} 0 = & \frac{1}{\langle f(k) \rangle} \left[-f(k)\mathrm{P}(k) + \theta(k) f(k-1)\mathrm{P}(k) \right] \\ & + \frac{1}{\bar{k}} \left[-\theta(k) k \mathrm{P}(k) + (k+1)\mathrm{P}(k) \right] , \end{aligned} \tag{4.18}$$

where $\theta(k)$ is the Heaviside step function and has been introduced to make the equation valid for all $k \geq 0$. This condition can be written in the form

$$0 = J(k+1) - J(k)\theta(k) , \tag{4.19}$$

with

$$J(k) = \frac{k}{\bar{k}}\mathrm{P}(k) - \frac{f(k-1)}{\langle f(k) \rangle}\mathrm{P}(k) . \tag{4.20}$$

Now, by iteration, it is found that

$$J(k) = 0 , \tag{4.21}$$

which implies

$$\mathrm{P}(k+1) = \frac{\bar{k} f(k)}{(k+1)\langle f(k) \rangle}\mathrm{P}(k) . \tag{4.22}$$

Equation (4.22) can be solved recursively. Note that if the choice $\bar{k} = \langle f(k) \rangle$ is made—which amounts to choosing a particular density of links in the system, for a given preference function—then the degree distribution is given by

$$\mathrm{P}(k) = \left[\prod_{m=0}^{k-1} \frac{f(m)}{m+1} \right] \mathrm{P}(0) . \tag{4.23}$$

Thus, if the preference function behaves asymptotically as

$$f(k) \sim k + 1 - \gamma + \mathcal{O}\left(\frac{1}{k}\right) , \tag{4.24}$$

where γ is some finite number, then one finds a scale-free degree distribution of the system with exponent γ

$$\mathrm{P}(k) \sim k^{-\gamma} . \tag{4.25}$$

If $\bar{k}$ is chosen below the particular density, then it is found that the degree distribution has an exponential form; if chosen above then the degree distribution has a power-law form, but with an additional peak indicating the presence of a condensation of links in the system. This is identical behaviour to that seen in a zero-range process with an appropriate choice of hop-rates, see Chapter 2 Section 2.3—the connection between the zero-range process and rewiring network models is discussed in depth in Section 4.7.

The above construction is referred to by Dorogovtsev, Mendes and Samukhin as a Canonical Ensemble of networks. Microcanonical and grand canonical formulations also exist [139, 8]. Though the ensembles are given names from standard statistical mechanics, the network ensembles do not necessarily correspond exactly to their namesakes. The microcanonical ensemble has the number of nodes and links fixed, but also the degree sequence, i. e. a specification of the number of nodes with each possible degree, k. Thus the microcanonical ensemble is constructed by assigning an equal probability to each network that has a given degree sequence and numbers of nodes and links. The grand canonical ensemble on the other hand, is constructed by randomly deleting links at some rate λN and having edges appear between nodes i and j with rate $f(k_i)f(k_j)$, where λ is a constant and $f(k)$ is the same preference function as for the canonical ensemble. If the parameters of each of the ensembles are chosen appropriately, then they are all equivalent in the thermodynamic limit.

The steady state of the canonical ensemble defined above is known to share the same partition function as the balls-in-boxes model [124, 125]. The partition functions defined through minifield theories in [123, 126, 130] were also noted to be connected to the same balls-in-boxes model. This model was originally defined statically, but there exist many dynamic models that share the same partition function, a particular one, the zero-range process, will be discussed in the following section.

4.7 The zero-range process as a network model

In the previous section it was detailed how several of the equilibrium statistical ensembles of networks studied enjoyed a connection with the balls-in-boxes model. The balls-in-boxes model is a model that involves distributing a number of balls among a number of boxes. All these cases shared the property that the partition function took the form of a sum over a quantity that factored over nodes, each factor depended only on the degree of the relevant node; in other words, the Hamiltonian was a function of node degree and additive over nodes. Indeed these ensembles shared a partition function with the balls-in-boxes model and so of course the partition function of the balls-in-boxes model is a sum over a quantity that factors over boxes and each factor depends only on the number of balls in the relevant box. This leads to the obvious network interpretation of a configuration of a number of balls distributed among a number of boxes:

- Boxes are interpreted as nodes of the network.
- The number of balls in a box is interpreted as the degree of the node, i. e. balls

are interpreted as link ends.

In the language of the zero-range process, boxes would be sites and balls would be particles; sites are interpreted as nodes and particles as link ends. In fact any model that involves the static placement or dynamic motion of particles on a lattice where the sites can be labelled can be regarded, on some level, as a model of a network as its configurations can be mapped to degree distributions.

For the mapping to be sound, it is required that there be an even number of particles in the system: each link contributes twice to the degree sequence. However, there are ways around this constraint. For example, one could consider a directed network and have two types of particles of equal number, representing in and out links. Having an equal number both types guarantees an even number of link ends overall and the directedness of the network can be ignored in calculations, if desired.

The above mapping between interacting particle systems and networks also only gives direct information about the degree sequence of the network. A given configuration of the particle system defines the degree of each node, the degree sequence, but this degree sequence may, in general, correspond to multiple distinct network configurations as depicted in Figure 4.6. If the particles could be paired and these pairs distinguished and tracked, then the evolution of the particle system would give the explicit evolution of the networks structure. The question of how, and if, the behaviour of the hypothetical paired model can be recreated from the indistinguishable particle model at an ensemble level is discussed later.

The ZRP in particular forms a model of a dynamically evolving undirected network. Links rewire from nodes with a rate that depends on the degree of the node at the end of the link that is to be rewired, and rewire to a node chosen at random from all others (in the fully connected, homogeneous version of the ZRP). It has been shown in Chapter 2 that the ZRP can undergo a condensation transition whereby a finite fraction of the total number of particles condenses onto a single site, in the network context this would correspond to a single node capturing a finite fraction of the available links. It was also shown that at criticality the ZRP can have a power-law occupancy distribution, corresponding to a scale-free degree distribution from a networks perspective. Thus the ZRP, as a network model, shares properties with several of the models discussed above, i. e. it can have a power-law degree distribution at a critical point and it can show condensation transitions in the degree. The connection can be elucidated further for some of these models, in particular the connection to the equilibrium canonical-ensemble rewiring network model of Dorogovtsev, Mendes and Samukhin (discussed in Section 4.6.1) is very close: with appropriate choices of

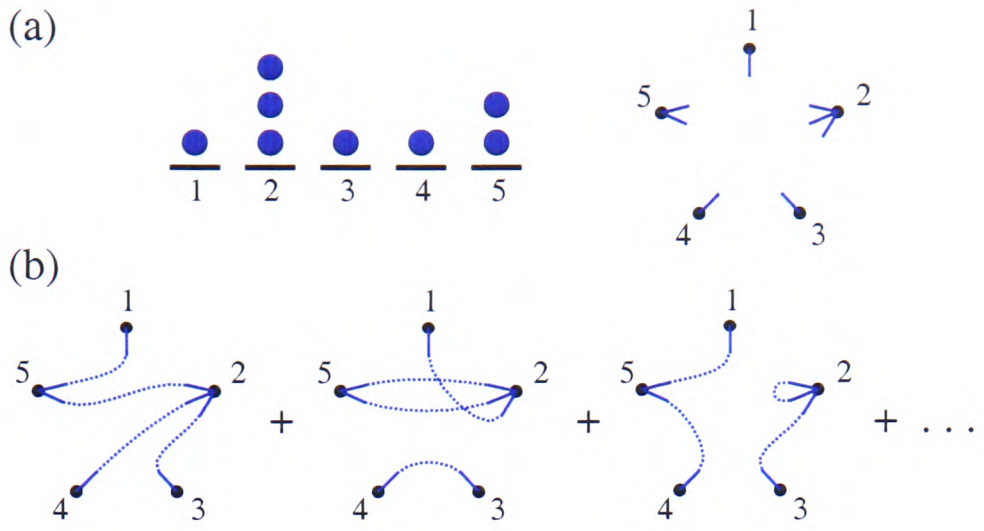


Figure 4.6: A given degree sequence may in general correspond to multiple distinct network configurations. (a) A particle configuration and the corresponding stub configuration that is the starting point for all the possible pairings of particles to create links. (b) Some of the possible distinct network configurations found by connecting the links, including repeated links, self links and disconnected components.

rates the two share the same degree distribution and partition function. In the case of the ZRP a node would be chosen at random and then a link attached to this node would be chosen, also at random. This link would then be rewired away from the chosen node to a second randomly chosen node with a probability that depends on the degree of the first node. At first site this seems quite different to the model discussed in Section 4.6.1: in this case links are chosen at random and rewired to a node with a probability that depends on the degree of this target node. To see that the two models can share the same degree distribution the Master equation approach (4.16-4.23) can be followed for the ZRP

$$\begin{aligned}\frac{\partial \mathrm{P}(k,t)}{\partial t} = &- \theta(k)u(k)\mathrm{P}(k,t) + u(k+1)\mathrm{P}(k+1,t) \\ &- \langle u(k)\rangle \,\mathrm{P}(k) + \theta(k)\,\langle u(k)\rangle \,\mathrm{P}(k-1)\;,\end{aligned} \tag{4.26}$$

where k denotes the number of particles on a site of the ZRP, or equivalently the degree of the node and $\langle u(k)\rangle$ is the ensemble average of the hop rate, $u(k)$. In the steady state the degree distribution will not depend on time and so, as before, the relation

$$0 = -\theta(k)u(k)\mathrm{P}(k,t)+u(k+1)\mathrm{P}(k+1,t)-\langle u(k)\rangle \,\mathrm{P}(k)+\theta(k)\,\langle u(k)\rangle \,\mathrm{P}(k-1)\;, \tag{4.27}$$

is found, with

$$J(k) = u(k)P(k) - \langle u(k) \rangle P(k-1) \, . \tag{4.28}$$

Again, by iteration, it is found that $J(k) = 0$ and thus the relation

$$\mathrm{P}(k) = \frac{\langle u(k) \rangle}{u(k)} \mathrm{P}(k-1) \, , \tag{4.29}$$

is found. Equation (4.29) can be solved recursively. If the choice $\langle u(k) \rangle = 1$ is made—this amounts to choosing the number of particles, or links, in the system—then the degree distribution is given by

$$\mathrm{P}(k) = \left[\prod_{m=1}^{k} \frac{1}{u(k)} \right] \mathrm{P}(0) \, . \tag{4.30}$$

Thus, if the further choice

$$u(k) = k/f(k-1) \, , \tag{4.31}$$

is made, where $f(k)$ is the attachment probability of the canonical equilibrium rewiring network of Dorogovtsev, Mendes and Samukhin, then the ZRP will share exactly the same degree distribution (4.23 as this model. Furthermore, by condition (4.24), if $u(k)$ behaves as

$$u(k) \sim 1 + \frac{\gamma}{k} + \mathcal{O}\left(\frac{1}{k^2}\right) \, , \tag{4.32}$$

asymptotically, then $\mathrm{P}(k)$ will have a power-law tail, i. e. the network will have a scale-free degree distribution

$$\mathrm{P}(k) \sim k^{-\gamma} \, . \tag{4.33}$$

Note that this Master-equation approach gives the probability that a site contains a certain number of particles—or the degree distribution—in a grand-canonical form with $\langle u(k) \rangle$ playing the role of the fugacity. It was shown in Chapter 2 Section 2.4 that the fugacity is equal to the average hopping rate in a general ZRP. This motivates the assertion that the choice $\langle u(k) \rangle = 1$ is the same as choosing the number of particles in the system. Thus, as with the grand-canonical ensemble, this treatment will break down at a critical particle or link number, signalling a phase transition.

Thus far it has been shown that the degree distributions of the ZRP and the Dorogovtsev-Mendes-Samukhin canonical rewiring model are equal for an appropriate choice of parameters in each case. However, as discussed briefly above, a given degree distribution or sequence in general can correspond to multiple network configurations. Only by labelling pairs of particles as belonging to single links could a distinct network be tracked through the dynamical evolution of a ZRP. Since the interest is in calculating ensemble averages of properties, if the ensemble average for the paired

particle system can be reconstructed from the indistinguishable particle system, then ensemble averages of network properties that do not depend solely on the degree can be correctly obtained from the model that only keeps track of the degree sequence. The obvious ansatz to make is that, for a given degree sequence, the statistical weight of a distinct network configuration is proportional to the number of ways that it can be created by pairing particles to create links at random. Thus if the particles are not labelled, then a form of ergodicity is assumed, i. e. that all possible pairings of particles will be explored equally. In this respect modelling a network in this way draws parallels with the configuration model, and its relatives, discussed in Section 4.4.1. This is a model that creates distinct networks from a given degree sequence—which may be drawn from some distribution—by assigning a number of stubs to each node in accordance with the degree, similar to the diagram shown in Figure 4.6 (a); these stubs are then randomly chosen in pairs and linked together. These models generally restrict to graphs without self-links or multiple connections, but for the ZRP this restriction is not always made.

When doing an ensemble average, the concept becomes quite similar to the random graphs with arbitrary degree distribution discussed in Section 4.4.1. Only the degree distribution is generated by the ZRP model—to calculate any distinct network properties the ensemble is averaged over network configurations that are maximally random given the degree distribution.

The ergodic assumption of random pairing of particles of the ZRP to give links of a network configuration can further be justified by looking at the connection with the canonical rewiring network of Dorogovtsev, Mendes and Samukhin. In Section 4.6.1 a Master Equation formalism that only followed the degree distribution was presented. However, in the original paper [139] a formalism that followed the actual configuration of the network during its evolution was presented.

This formalism can be followed for the ZRP network dynamics and this proves instructive in how individual networks should be tracked. The starting point is the assumption of a detailed balance in the space of undirected network configurations, i. e.

$$W(g', g)\mathrm{P}(g) = W(g, g')\mathrm{P}(g') \,, \tag{4.34}$$

where g, g' label distinct network configurations, $\mathrm{P}(g)$ is the steady state probability of a given network configuration, g, and $W(g', g)$ is the transition rate from a configuration, g, to another, g'. No restriction on multiple- and self-links is made. Now let g_{ij} be the number of links between nodes i and j and g_{ii} be twice the number of self-links to node i. In the ZRP network dynamics the transition rate for the rewiring

of a link (i,j) to (i,k) is given by

$$W(g',g) = \frac{1}{L^2}\frac{g_{ij}}{q_j}u(q_j)\ , \tag{4.35}$$

i. e. node j is chosen with probability $1/L$, a link (i,j) is then chosen with probability g_{ij}/q_j, this link is then rewired from node j with rate $u(q_j)$ and node k is chosen as the new connection for the link with probability $1/L$. It has also been assumed that the networks here labelled g and g' contain the relevant links. Thus the balance condition reads

$$\frac{1}{L^2}\frac{g_{ij}}{q_j}u(q_j)\mathrm{P}(g) = \frac{1}{L^2}\frac{g'_{ik}}{q'_k}u(q'_k)\mathrm{P}(g')\ , \tag{4.36}$$

where primed variables label those of the configuration g' and they are related to their unprimed counterparts as follows

$$\begin{aligned} q'_k &= q_k + 1 \qquad & g'_{ik} &= g_{ik} + 1 + \delta_{ik} \\ q'_j &= q_j - 1 \qquad & g'_{ij} &= g_{ij} - 1 - \delta_{ij}\ . \end{aligned} \tag{4.37}$$

The Kronecker deltas account for the fact that g_{ii} is defined as twice the number of self-links, so when one of these is added or removed g_{ii} changes by two. The balance equation (4.36) can be solved with a solution of the form

$$\mathrm{P}(g) \propto \prod_{i=1}^{N} p(q_i)\chi_d(g_{ii}) \prod_{j<k=1}^{N} \chi(g_{jk})\ , \tag{4.38}$$

where $p(q)$, $\chi_d(g)$ and $\chi(g)$ are functions to be determined. Inserting (4.38) into (4.36), taking $i \neq j \neq k$, cancelling common factors and rearranging a little yields

$$\frac{g_{ij}\chi(g_{ij})}{\chi(g_{ij}-1)}\frac{u(q_j)p(q_j)}{q_j p(q_j-1)} = \frac{(g_{ik}+1)\chi(g_{ik}+1)}{\chi(g_{ik})}\frac{u(q_k+1)p(q_k+1)}{(q_k+1)p(q_k)} = \text{constant}\ , \tag{4.39}$$

where the constant can be taken to be equal to one without loss of generality. This in turn yields

$$p(q_i+1) = \frac{(q_i+1)p(q_i)}{u(q_i+1)} \tag{4.40}$$

$$\chi(g_{jk}+1) = \frac{\chi(g_{jk})}{g_{jk}}\ . \tag{4.41}$$

Taking $i = j$, cancelling common factors and rearranging yields

$$\frac{g_{ii}\chi_d(g_{ii})}{\chi_d(g_{ii}-2)}\frac{u(q_i)p(q_i)}{q_i p(q_i-1)} = \frac{(g_{ik}+1)\chi(g_{ik})+1}{\chi(g_{ik})}\frac{u(q_k+1)p(q_k+1)}{(q_k+1)p(q_k)} = 1\ , \tag{4.42}$$

where again the constant can be taken equal to one without loss of generality. This then yields

$$\chi_d(g_i i) = \frac{\chi_d(g_i i-2)}{g_i i}\ , \tag{4.43}$$

and choosing $i = k$ gives the same condition. The equations (4.40), (4.41) and (4.43) can be solved recursively, giving

$$p(q_i) = q_i! \prod_{m=1}^{q_i} \frac{1}{u(m)}, \qquad p(0) = 1 \tag{4.44}$$

$$\chi(g_{jk}) = \frac{1}{g_{jk}!} \tag{4.45}$$

$$\chi_d(g_{ii}) = \frac{1}{g_{ii}!!}, \tag{4.46}$$

where $g_{ii}!! = g_{ii}.(g_{ii}-2)\ldots 4.2$, for $g_{ii} > 0$ and g_{ii} even, is the usual double factorial. Note that $p(q_i)$ is equal to $f(q_i)q_i!$, where $f(q_i)$ is the usual unnormalised probability factor from the basic ZRP (2.3). Thus the probability of a given network configuration, g, in the steady state is given by

$$\mathrm{P}(g) \propto \prod_{i=1}^{N} f(q_i) \frac{q_i!}{g_{ii}!!} \prod_{j<k=1}^{N} \frac{1}{g_{jk}!}. \tag{4.47}$$

Note that this is a product of the probability of obtaining the degree sequence $\{q_i\}$ ($\propto \prod_i f(q_i)$ (2.2)) and the number of ways of obtaining the given network configuration g by randomly pairing particles to make links given by the factorials of the above equation, as has been shown in [139] Appendix A. Thus, for a given degree sequence, averaging over all possible pairings of particles will correctly calculate, for the given dynamics, any network quantities even if they do not depend solely on the degree.

Having a rewiring networks model in the language of the ZRP carries several immediate advantages. The steady state for the ZRP is known for the basic model and for many generalisations, see Chapter 2. For example the inhomogeneous version of the ZRP could be used to implement fitness into the model, the probability of a degree sequence is then given by (2.2). Also the ZRP on an arbitrary lattice could be used to put spatial information in the network, i. e. links could be only allowed to rewire to a node that is local to the node which they are rewiring from. The ZRP can also be solved for various nonequilibrium choices of the dynamics.

In this section the validity of the ZRP as a network model has been discussed along with the method for correctly recreating average network properties given the fact that the ZRP itself only keeps track of the degree of the network, not any distinct network configurations. The discussion given above was of the most basic use of the ZRP as a network model. It is worth noting that a slightly different model exists in the literature [141], where a specific prescription for a bipartite network involving the ZRP allowed the giant component to be studied in the unipartite projection of the bipartite network. It should also be noted that due to its property of factorisation

for quite general lattices, see Chapter 2 Section 2.5.1, the ZRP is also suitable as a model of processes happening on networks [142, 143]. In the following chapters further original applications of ZRPs as models networks will be discussed.

4.8 Summary

In this chapter the concept of a network (loosely speaking a collection of nodes connected by links) has been introduced and some of the characteristics that are used to classify network behaviour were given. Several of the seminal models of these complex networks were discussed in detail. Some of the more recent models that directly use the concepts of statistical mechanics were also discussed, culminating in a discussion of use of the ZRP as a model of networks. It was shown that the configuration of the ZRP could be mapped onto a degree distribution of a network and that from this an ensemble of network configurations could be generated, for each of which a statistical weight can be determined.

Chapter 5

Creation and Annihilation of Particles/Links: Evaporating the Condensate

5.1 Introduction

It was seen in the previous chapter that a great many complex networks have scale-free or power-law degree distributions. It was also seen that several of the statistical ensemble models of networks only possessed such a distribution at a critical point. Thus, it would be advantageous if some ingredient could be added to these models that allowed them to exhibit critical behaviour without fine-tuning the parameters such that the system is at criticality. The nature of this added ingredient would then give another mechanism for generating networks with scale-free degree distributions and perhaps further insight into how and why so many real networks attain scale-free states.

The ZRP model of a network, as discussed in Chapter 4 Section 4.7, was one of the models that had a power-law degree distribution only with a critical density of links in the system. Below the critical density the degree distribution was exponentially decaying and the system was in a low density, fluid state. Above the critical density the degree distribution took the form of a power-law distribution with an additional peak at high degree and the system was in a condensed state. This condensed state consists of a single node with a finite fraction of all the links and with the rest of the nodes having degrees according to a power-law distribution. An obvious way to generate scale-free degree distributions therefore is to add an element to the dynamics of the model that suppresses the condensate in a way that leaves the power-law background

intact. In this chapter it is shown how the introduction of non-conservation in an appropriate fashion causes the ZRP network model to give critical behaviour in an entire region of the parameter space rather than just at a critical point.

Of course obtaining a critical phase by adding creation and annihilation is behaviour that is not restricted to the network model; it also applies to the ZRP in its own right. A ZRP with critical behaviour in a region of the parameter space could find application in areas other than networks. It is in the language of the ZRP that the non-conservation rules will first be discussed and analysed; both because the language of the ZRP is more general than the language of networks and because the non-conservation actually introduces some subtleties in the network version of the model which makes the analysis slightly more involved.

5.2 ZRP with creation and annihilation of particles

5.2.1 Introduction

In this section a generalised ZRP with the additional elements of creation and annihilation of particles is studied. These elements are realised in a more specific way than say a grand-canonical ensemble and allow a critical phase in the model.

5.2.2 Model definition

For simplicity a specific model is studied here; possible generalisations are discussed in Section 5.2.5. A ZRP with L sites and N particles is considered; the number of sites is conserved, but the number of particles is allowed to vary under the dynamics. The particles hop from site to site with rates $u(n_\mu)$, which depends on the number of particles at the departure site, n_μ, as in the usual ZRP. The lattice considered is fully connected, i. e. a particle that hops from a site will arrive at any other site with equal probability. In addition to this hopping process, two further processes happen concurrently: particles annihilate from site μ with rate $a(n_\mu)$, which depends on the number of particles at the site, n_μ and particles are created at a site with rate c, which for simplicity is taken to be independent of the number of particles at the site. The following explicit rates are studied:

- **Hop rate**

$$u(n) = \left(1 + \frac{b}{n}\right)\theta(n)\,, \tag{5.1}$$

where $b > 0$ is a constant and $\theta(n)$ is the usual Heaviside step function [10].

- **Annihilation rate**

$$a(n) = \left(\frac{n}{L}\right)^k , \tag{5.2}$$

  where $k > 0$ is a constant.

- **Creation rate**

$$c = \frac{1}{L^s} , \tag{5.3}$$

  where $s > 0$ is a constant.

The above rates are chosen for their simplicity and because they contain the basic elements expected to give the desired behaviour of a power-law occupancy distribution, namely a hop rate that gives a power-law distribution at a critical point and an annihilation rate that increases with particle number, thus which most favourably annihilates from condensed sites. The creation rate at a site c (5.3) is independent of the number of particles at the site but depends on the system size. The overall creation rate in the system is given by $L^{(1-s)}$ and can be controlled by varying s. This is important as the creation rate can be thought of as the driving rate for the system. The annihilation rate $a(n)$ (5.2) depends on both the size of the system and the number of particles at a site. By increasing k the annihilation is made to happen more from highly occupied sites. The case $k = 1$ corresponds to the independent annihilation of particles, thus the form of the annihilation rate allows for easy comparison with this case. The hop rate $u(n)$ (5.1) is one of the simplest that is known to give a power-law distribution at the critical point of a conserving ZRP as discussed in Chapter 2 at the end of Section 2.3. Varying the parameter b allows the exponent of this distribution to be changed. Note that as $b > 2$ is required for a condensation transition and hence for there to be a critical point, the remaining discussion is restricted to the case $b > 2$.

5.2.3 Analysis

It was seen in Chapter 2 Section 2.2.2 that the ZRP possesses a factorised steady state, i. e. the probability of a given configuration, $\mathrm{P}(n_1, n_2, \ldots, n_L)$, takes the form of a product over sites of quantities that depend only on the number of particles on each site. Thus, in a homogeneous system it is sufficient to study the probability of the occupancy of a single site to gain full knowledge of the system. In contrast, the generalised ZRP (with creation and annihilation) studied in this chapter is, in general, not expected to have a factorised steady state. However, the approximation of a factorised steady state, $\mathrm{P}(n_1, n_2, \ldots, n_L) \propto \prod_{i=1}^{L} \mathrm{p}(n_i)$, where $\mathrm{p}(n)$ is the probability that a single site has occupancy n, is made for the following analysis. For the fully connected geometry considered, such an approximation is expected to become exact in

the limit $L \to \infty$. Within this approximation the master equation for the probability of occupancy of each site, $\mathrm{p}(n)$ is given by

$$\begin{aligned}\frac{\partial \mathrm{p}(n)}{\partial t} &= [u(n+1)+a(n+1)]\,\mathrm{p}(n+1) - [v+c]\,\mathrm{p}(n) \\ &\quad - \{[u(n)+a(n)]\,\mathrm{p}(n) - [v+c]\,\mathrm{p}(n-1)\}\,\theta(n)\,.\end{aligned} \tag{5.4}$$

The terms on the right-hand side of this equation represent the probability currents in and out of a state with n particles on a site. Grouping them into the terms within the square brackets and including the multiplicative probabilities, they are explained as follows: the first term represents the current in due to a hop out of or annihilation from a site with $n+1$ particles; the second term represents the current out due to a hop into or a creation on a site with n particles; the third term represents the current out due to a hop out of or an annihilation from a site with n particles; finally, the fourth term represents the current in due to a hop into or creation at a site with $n-1$ particles. The theta function in the second line ensures that (5.4) correctly includes the boundary case $n = 0$. The current due to a hop into a site contains v, the average hop rate from a site, which depends on the state of the system and is given by

$$v = \sum_{n=1}^{\infty} u(n)\mathrm{p}(n)\,. \tag{5.5}$$

Thus the master equation (5.4) is not as simple as it might first seem as it contains more $\mathrm{p}(n)$ dependence than is explicitly shown.

In the steady state, $\mathrm{p}(n)$ is stationary in time and (5.4) becomes

$$\begin{aligned}0 &= [u(n+1)+a(n+1)]\,\mathrm{p}(n+1) - [v+c]\,\mathrm{p}(n) \\ &\quad - \{[u(n)+a(n)]\,\mathrm{p}(n) - [v+c]\,\mathrm{p}(n-1)\}\,\theta(n)\,.\end{aligned} \tag{5.6}$$

This can be written in the form

$$0 = J(n+1) - J(n)\theta(n)\,, \tag{5.7}$$

with

$$J(n) = [u(n)+a(n)]\,\mathrm{p}(n) - [v+c]\,\mathrm{p}(n-1)\,. \tag{5.8}$$

Thus, by iteration it is found

$$J(n) = 0\,, \tag{5.9}$$

which implies

$$\mathrm{p}(n) = \frac{v+c}{u(n)+a(n)}\,\mathrm{p}(n-1)\,, \tag{5.10}$$

which in turn is solved iteratively to give

$$\mathrm{p}(n) = \frac{(v+c)^n}{\prod_{m=1}^{n}[u(m)+a(m)]}\,\mathrm{p}(0)\,. \tag{5.11}$$

Then to find p(n) explicitly, v and p(0) must be found. The solution must also satisfy the following constraint equations: the probability must be normalised

$$\sum_{n=0}^{\infty} \mathrm{p}(n) = 1 \; ; \tag{5.12}$$

and in the steady state the overall creation and annihilation rates must balance

$$\sum_{n=1}^{\infty} a(n)\mathrm{p}(n) = c \; . \tag{5.13}$$

Note that only two of (5.5), (5.12) and (5.13) are independent.

The behaviour of the system is determined by finding the large L asymptotic behaviours of v and p(n) that satisfy (5.5) and (5.13). The system defined by the dynamics (5.1), (5.2) and (5.3) has several phases depending on the values of the parameters b, k and s. The phase diagram is shown in Figure 5.2.3. In what follows the three phases, low density, critical and high density, are determined and analysed.

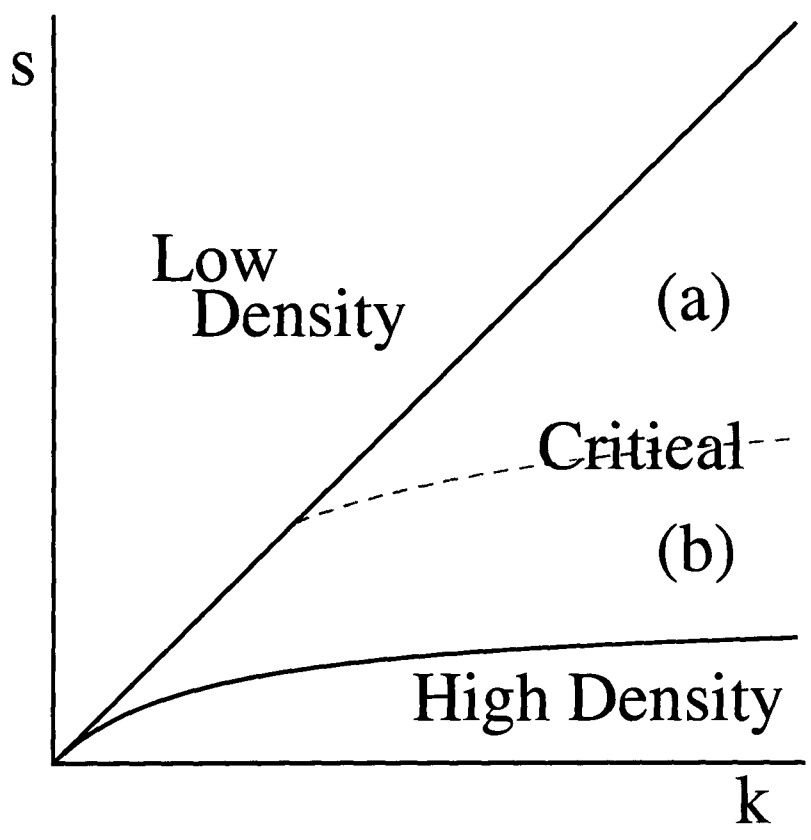


Figure 5.1: Typical phase diagram for both the non-conserving ZRP model and the non-conserving network model with self links allowed. Shown in the k–s plane with b fixed. The parameters k, s, b are defined in (5.1), (5.2), (5.3), (5.63), (5.64) and (5.65).

5.2.4 Phase analysis

The starting point of the analysis is the creation-annihilation balance equation (5.13). Rearranging a little and inserting the expressions for p(n) (5.11), c (5.3) and $a(n)$

(5.2) gives

$$\begin{aligned} L^{k-s} &= \sum_{n=1}^{\infty} n^k \mathrm{p}(n) \\ &= \mathrm{p}(0) \sum_{n=1}^{\infty} n^k \exp\left[n \ln(v+c) - \sum_{m=1}^{n} \ln(u(m)+a(m)) \right] . \end{aligned} \tag{5.14}$$

The form of the left hand side of (5.14) identifies two regions of s (with k fixed): a region ($s > k$) with the LHS going to zero as some negative power of L and a region ($s < k$) with the LHS diverging as some positive power of L. These two regions require different treatments.

Using (5.14) the following regions of behaviour are identified

- **Low density phase** ($s > k$): In this region the probability distribution is exponentially decaying and the global density of particles tends to zero in the thermodynamic limit.

- **Critical phase** ($k/(k+1) < s < k$): In this region the probability distribution has a power-law form in the thermodynamic limit. The large but finite behaviour of the probability distribution can be used to classify two sub-phases: critical sub-phase (a) ($kb/(k+1) < s < k$), where the probability distribution is a power-law with an exponential cut-off; and critical sub-phase (b) ($k/(k+1) < s < kb/(k+1)$), where the probability distribution is a power-law with a weak peak at high n.

- **High density phase** ($s < k/(k+1)$): In this region the probability distribution tends to a delta peak and all the sites have large occupancies.

In the following these phases will be analysed in detail.

Low density phase: $s > k$

If the parameter s is greater than the parameter k, then the LHS of (5.14) will go to zero in the limit of large system size. Thus to satisfy (5.14), $\mathrm{p}(n)$ is a rapidly decreasing function of n. Note that $\mathrm{p}(0)$ plays no part in the overall rate of annihilation and so can can be of any size to satisfy the normalisation condition 5.12. Assuming that $\mathrm{p}(n)$ is a monotonically decreasing function of n and that $\mathrm{p}(1)$ dominates the sum on the RHS of 5.14, then $\mathrm{p}(1) \simeq L^{k-s}$. Furthermore, if $\mathrm{p}(1) \gg \mathrm{p}(2)$, then $\mathrm{p}(1) \simeq \langle N \rangle / L$, where $\langle N \rangle / L$ is the average number of particles in the system. Thus $\langle N/L \rangle \sim L^{k-s}$, i. e. the average density of the system will tend to zero as the system size tends to infinity. This is consistent with having a very small v, i. e. a small average current from

a site due to a low density of particles in the system. Strictly $v+c$ is small such that $v+c \simeq v \sim L^{k-s}$, to leading order in L. Then $\mathrm{p}(n) \sim L^{-(s-k)n}$, with $\mathrm{p}(1) \sim L^{k-s}$. The mean number of particles behaves as L^{k-s+1} and increases sub-linearly with system size.

For large but finite systems it is possible to get an approximate theory form for the probability distribution in the low density phase. This is achieved by assuming that the balance equation is dominated by $\mathrm{p}(1)$ and thus v is given by L^{k-s}. This is then inserted back into (5.14) to give the asymptotic theory line

$$\mathrm{p}(n) \sim L^{n(k-s)} \,. \tag{5.15}$$

This equation can then be compared with results from simulations.

Analysis of region: $s < k$

If the parameter s is less than the parameter k, then the LHS of (5.14) will diverge as L^{k-s}. Thus, the sum on the RHS must also diverge in this way and it is dominated by terms at large n. Expanding $\sum_{m=1}^{n} \ln(u(m)+a(m))$ for large n yields

$$\sum_{m=1}^{n} \ln(u(m)+a(m)) \simeq \int_1^n \mathrm{d}m \, \ln\left[1+\frac{b}{m}+\left(\frac{n}{L}\right)^k\right] \simeq b\ln n + \frac{n^{k+1}}{(k+1)L^k} \,. \tag{5.16}$$

This gives a decreasing part to the RHS of (5.14) with a power-law factor n^{-b} and an exponentially decreasing factor $\exp(-n^{k+1}/((k+1)L^k)$. Thus equation (5.14) can only be satisfied if $v+c$ approaches one as L tends to infinity: if $v+c$ is too small, $\exp(n\ln(v+c))$ will give an exponentially decreasing factor that is sufficiently rapid in its decrease to cause the RHS to converge; if too big, a rapidly increasing exponential factor will be generated which will cause the RHS to diverge too strongly. Thus, in this region, $v+c$ is taken to have the form

$$v+c \simeq 1+g(L) \,, \tag{5.17}$$

where $g(L)$ tends to zero as L tends to infinity. Hence, to leading order in L,

$$\ln(v+c) \simeq g(L) \,. \tag{5.18}$$

Inserting (5.16) and (5.18) into (5.14) yields the asymptotic forms for the site occupancy distribution

$$\mathrm{p}(n) \sim n^{-b} \exp\left[n\,g(L) - \frac{n^{k+1}}{(k+1)L^k}\right] \,, \tag{5.19}$$

and the creation-annihilation balance equation

$$L^{k-s} \sim \int_1^\infty \mathrm{d}n \, n^{k-b} \exp\left[n\,g(L) - \frac{n^{k+1}}{(k+1)L^k}\right] \,, \tag{5.20}$$

where the sum has been approximated by an integral. Equation (5.20) is key to the analysis of all the remaining phases: the remaining regions can be identified by determining the large L asymptotic behaviour of $g(L)$ that satisfies (5.20). Inserting this into (5.19) gives the form of the site occupancy distribution, $\mathrm{p}(n)$, via which the nature of the phases can be identified.

High density phase: $s < k/(k+1)$

The low density phase is one of the extremes of the system, where $\ln(v+c)$ is less than one asymptotically and so the occupancy distribution is rapidly decreasing. Another extreme of the system will be when $v+c$ approaches one from above asymptotically with a form that is the largest possible without causing the RHS of (5.20) to diverge too rapidly. This is when $g(L)$ behaves as some inverse power of L:

$$g(L) \simeq \frac{g_h}{L^y} \,, \tag{5.21}$$

where g_h and y are constants to be determined. Inserting this form into (5.20) yields

$$L^{k-s} \sim \int_1^\infty \mathrm{d}n\, n^{k-b} \exp\left[g_h \frac{n}{L^y} - \frac{n^{k+1}}{(k+1)L^k} \right] \,. \tag{5.22}$$

It is then assumed that the above integral will be dominated by the maximum of the exponential. Calling the exponent $\psi(n)$, it is found:

$$\psi(n) = g_h \frac{n}{L^y} - \frac{n^{k+1}}{(k+1)L^k} \tag{5.23}$$

$$\psi'(n) = \frac{g_h}{L^y} - \frac{n^k}{L^k} \tag{5.24}$$

$$\psi''(n) = -k\frac{n^{k-1}}{L^k} \,. \tag{5.25}$$

The maximum is given by the point $n^\star$

$$\psi'(n^\star) = 0 \tag{5.26}$$

$$n^\star = g_h^{1/k} L^{(k-y)/k} \sim L^{(k-y)/k} \,. \tag{5.27}$$

At this maximum point, the exponent takes the value

$$\psi(n^\star) = L^{1-y(1+1/k)} \frac{k\, g_h^{(k+1)/k}}{(k+1)} \sim L^{1-y(1+1/k)} \,. \tag{5.28}$$

Thus when $y < k/(k+1)$ there is a large peak in the unnormalised occupancy distribution, with a height behaving as an exponential of a positive power of L. This peak is thus very high and it is expected to dominate the $\mathrm{p}(n)$ distribution for large L.

In the following the dominance of the peak over the distribution is motivated. By Taylor expanding $\psi(n)$ around its maximum, the exponential can be treated as a Gaussian and thus the peak has a width $2\,|\psi''(n^\star)|^{-1/2}$. Now,

$$\psi''(n^\star) = -k g_h^{(k-1)/k} L^{-1-y(1-1/k)} \sim -L^{-1-y(1-1/k)} , \tag{5.29}$$

and the width behaves as

$$\text{width} \sim L^{(1/2)(1+y(1-1/k))} . \tag{5.30}$$

This is narrow on the scale of the mean, $n^\star \sim L^{(k-y)/k}$, as $L \to \infty$ and with $y < k/(k+1)$. Furthermore the peak will have an unnormalised weight (approximated by width $\times$ height)

$$\text{weight} \sim \exp\left[L^{1-y(1+1/k)}\right] L^{1/2+(y/2)(1-1/k)-b+yb/k} . \tag{5.31}$$

Thus it is clear that this peak dominates the $\mathrm{p}(n)$ distribution, and $\mathrm{p}(n)$ can be treated as a delta function centred on $n^\star$ for the asymptotic analysis.

Treating $\mathrm{p}(n)$ as a delta function centred on $n^\star$ gives the creation-annihilation balance equation

$$L^{k-s} \sim (n^\star)^k \sim L^{k-y} , \tag{5.32}$$

hence $y = s$. Recall that only for $y < k/(k+1)$ was the width of the peak narrow on the scale of the maximum location. Thus the above analysis is only valid in the region $s < k/(k+1)$. This is the high density phase, where $\mathrm{p}(n)$ is sharply peaked at $n^\star \sim L^{(k-s)/k}$ giving all sites an average number of particles that is an positive power of L and a super-extensive average total number of particles, $\langle N \rangle \sim L^{1-s-s/k}$.

For large but finite system sizes, the value obtained for y can be inserted into (5.22) to give a theory curve that can be compared with numerical simulations of the system

$$\mathrm{p}(n) \sim n^{-b} \exp\left[g_h \frac{n}{L^s} - \frac{n^{k+1}}{(k+1)L^k}\right] , \tag{5.33}$$

where g_h is an undetermined parameter that does not make a great difference to the shape of the line.

Critical phase: $k/(k+1) < s < k$

In between the low and high density phases a critical phase is found, where in the $L \to \infty$ limit the occupancy distribution $\mathrm{p}(n)$ approaches a power law. This phase can further be divided into two sub-phases, distinguished by the large but finite L forms of $\mathrm{p}(n)$.

Critical sub-phase (a): $kb/(k+1) < s < k$

When $s > k$ the system is in the low density phase and v is small giving a rapid cut-off. In critical sub-phase (a), the region where s first becomes less than k, v approaches 1 from below giving a cut-off that is slow enough that it leaves an appreciable region of power-law behaviour.

The case of v approaching one from below is satisfied when $g(L)$ is small and negative. The form that $g(L)$ takes for this region is

$$g(L) \simeq -\frac{g_{ca}}{L^x} \,, \tag{5.34}$$

where g_{ca} and x are parameters to be determined and it has been assumed that $x < s$ so that the effect of $g(L)$ is dominant, i. e. $\ln(v+c) \simeq -g_{ca}/L^x$ to leading order in L. Inserting this into (5.20) yields the balance equation

$$L^{k-s} \sim \int_1^\infty \mathrm{d}n\, n^{k-b} \exp\left[-g_{ca}\frac{n}{L^x} - \frac{n^{k+1}}{(k+1)L^k}\right] . \tag{5.35}$$

The second exponential cut-off in the balance equation above is fixed by the parameter k and if it were the dominant cut-off, then the balance equation could only be satisfied for a single value of s. Thus it is assumed that the first exponential cut-off dominates, i. e. $x < k/(k+1)$, this should be able to satisfy the balance equation as $g(L)$ can depend upon s. In what follows this dependence will be calculated.

Under the assumption that the first cut-off dominates, the integral on the RHS of (5.35) can be evaluated approximately by integrating the non cut-off part to an upper-limit determined by where the cut-off first becomes appreciably smaller than 1 giving

$$L^{k-s} \sim \int_1^{L^x/g_{ca}} \mathrm{d}n\, n^{k-b} \tag{5.36}$$

$$\sim L^{x(k-b+1)} \,, \tag{5.37}$$

where it has been assumed that $k > b-1$. If $k < b-1$ this phase cannot exist. For the balance to be satisfied it is required that

$$x = \frac{k-s}{k-b+1} \,. \tag{5.38}$$

Previously it was assumed that $x < k/(k+1)$ and $x < s$, for these to be consistent with (5.38) the following inequalities must be satisfied

$$s > \frac{kb}{k+1} \tag{5.39}$$

$$s > \frac{k}{k-b+2} \,, \tag{5.40}$$

with the first of these being a stronger condition than the second, if $k > b - 1$ as was also assumed. Thus in the region $k > s > kb/(k+1)$ the above analysis is consistent and the system is in a critical sub-phase with a $\mathrm{p}(n)$ distribution given by a power law with an exponential cut-off. The point at which this cut-off begins to appreciably affect the power-law behaviour goes to infinity with the system size as $L^{(k-s)/(k-b+1)}$.

The value calculated asymptotically for x can be inserted into the expression for $\mathrm{p}(n)$ to give a theory line

$$\mathrm{p}(n) \sim n^{-b} \exp\left[-g_{ca} \frac{n}{L^{(k-s)/(k-b+1)}}\right] , \tag{5.41}$$

this can then be compared with simulations of large but finite systems. The constant g_{ca} has not been determined by the analysis, but makes no difference to the shape of the curve and should have negligible effect for large enough systems.

Critical sub-phase (b): $k/(k+1) < s < kb/(k+1)$

In critical sub-phase (a) it was seen that a power-law with a cut-off form for $\mathrm{p}(n)$ could satisfy the balance equation, for $k > s > kb/(k+1)$. In the high density phase it was seen that a strong, high-n peak form for $\mathrm{p}(n)$ could satisfy the balance equation, for $s < k/(k+1)$. In between these regions the balance equation is satisfied by a $\mathrm{p}(n)$ that is a power-law with an additional weak peak at high n.

The form $g(L) \sim L^{-y}$ gave the high density phase, i. e. the unnormalised $\mathrm{p}(n)$ had a peak that was exponentially high in a positive power of L. In critical sub-phase (b), the unnormalised $\mathrm{p}(n)$ has a peak that is just a positive power of L in height, thus it dominates the balance equation without completely dominating the probability distribution. The form of $g(L)$ that gives this is

$$g(L) = \left(\frac{g_{cb} \ln L}{L}\right)^w , \tag{5.42}$$

where g_{cb} and w are parameters to be determined. Inserting this form into (5.20) yields the balance equation

$$L^{k-s} \sim \int_1^\infty \mathrm{d}n\, n^{k-b} \exp\left[n \left(\frac{g_{cb} \ln L}{L}\right)^w - \frac{n^{k+1}}{(k+1)L^k}\right] . \tag{5.43}$$

Writing $\mathrm{p}(n)$ as $\sim \exp(\psi(n)) n^{-b}$, it is found that

$$\psi(n) = n \left(\frac{g_{cb} \ln L}{L}\right)^w - \frac{n^{k+1}}{(k+1)L^k} , \tag{5.44}$$

$$\psi'(n) = \left(\frac{g_{cb} \ln L}{L}\right)^w - \frac{n^k}{L^k} , \tag{5.45}$$

$$\psi''(n) = -k \frac{n^{k-1}}{L^k} . \tag{5.46}$$

It is known from the analysis of critical sub-phase (a) that a power-law with any cut-off is not sufficient to satisfy the balance equation in the region considered here. Thus the weak peak of p(n) will dominate the contribution to the RHS of the balance equation(5.43). The most significant contribution from this peak will come from around the maximum of $\psi(n)$, which is given by the point $n^\star$

$$\psi'(n^\star) = 0 \tag{5.47}$$

$$n^\star = (g_{cb} \ln L)^{w/k} L^{(k-w)/k} , \tag{5.48}$$

and the height of the maximum of the peak will be

$$(n^\star)^{-b} \exp(\psi(n^\star)) . \tag{5.49}$$

Now $(n^\star)^{-b}$ clearly gives a power of L, albeit with a logarithmic correction, and $\psi(n^\star)$ is given by

$$\psi(n^\star) = (g_{cb} \ln L)^{w(k+1)/k} L^{(1-w(k/(k+1))} \frac{k}{k+1} . \tag{5.50}$$

As a positive power of L is required as the height of the weak peak, the above determines that $w = k/(k+1)$, hence

$$n^\star = g_{cb}^{1/(k+1)} L^{k/(k+1)} (\ln L)^{1/(k+1)} , \tag{5.51}$$

$$\psi(n^\star) = \frac{g_{cb}\, k}{k+1} \ln L = (\ln L)^{g_{cb}k/(k+1)} , \tag{5.52}$$

$$\psi''(n^\star) = -g_{cb}^{(k-1)/(k+1)} L^{-2k/(k+1)} (\ln L)^{(k-1)/(k+1)} . \tag{5.53}$$

Thus p(n) has a weak peak at $n^\star \sim L^{k/(k+1)}$ with width $2\,|\psi''(n^\star)|^{-1/2} \sim L^{k/(k+1)}$ and height $(n^\star)^{-b} \exp(\psi(n^\star)) \sim L^{k(g_{cb}-b)/(k+1)}$, ignoring logarithmic factors in all cases. Thus the weight ($\sim$ width $\times$ height) of the weak peak is asymptotically

$$\text{weight} \sim L^{(g_{cb}-b+1)k/(k+1)} . \tag{5.54}$$

The weak peak dominates the balance equation (5.43) and so, taking into account the fact that the peak is not a delta function in this case by including the weight, this becomes

$$L^{k-s} \sim (n^\star)^k \times \text{weight} \sim L^{(g_{cb}-b+1-k)k/(k+1)} , \tag{5.55}$$

which gives

$$g_{cb} = b - \frac{s(1+k)}{k} . \tag{5.56}$$

Now, for the peak to be weak, i. e. for it not to dominate the probability distribution, the weight of the peak must scale as a negative power of L. Inserting (5.56) into (5.54) gives weight $\sim L^{k/(k+1)-s}$ and $s > k/(k+1)$ is required to satisfy this, verifying the

lower boundary of the sub-phase. It is also required that the peak be strong enough that it can dominate over the contribution from the power-law part of the distribution in the annihilation part of the balance equation, i. e. over the largest possible contribution from a power-law with the largest cut-off, namely the annihilation from the RHS of (5.35) with the second cut-off dominant:

$$\int_1^\infty \mathrm{d}n\, n^{k-b} \exp\left[-\frac{n^{k+1}}{(k+1)L^k}\right] \sim \int_1^{L^{k/(k+1)}} \mathrm{d}n\, n^{k-b} \tag{5.57}$$

$$\sim \begin{cases} L^{k-kb/(k+1)} & \text{for } k > b-1 \\ \mathcal{O}(1) & \text{for } k < b-1\,. \end{cases} \tag{5.58}$$

The contribution from the weak peak to the annihilation is simply L^{k-s}, under the assumption that it dominates. Thus for the case $k > b-1$, the dominance of the weak peak in the balance equation is satisfied when $s < kb/(k+1)$—which borders on critical sub-phase (a). For $k < b-1$, critical sub-phase (a) does not exist and the analysis of critical sub-phase (b) extends to $s < k$—which borders on the low density phase.

Thus it has been proven that in critical sub-phase (b) the occupancy distribution, $\mathrm{p}(n)$, takes on the form of a power law with a weak peak at high n. The weight of this peak goes to zero as the system size tends to infinity, but in such a way that it dominates the balance equation and allows it to be satisfied. The values calculated for g_{cb} and w are inserted into the expression for $\mathrm{p}(n)$ to give a theory line

$$\mathrm{p}(n) \sim n^{-b} \exp\left[n\left\{\left(b - \frac{s(k+1)}{k}\right)\frac{\ln L}{L}\right\}^{k/(k+1)} - \frac{n^{k+1}}{(k+1)L^k}\right], \tag{5.59}$$

which can be compared with simulations.

Subdivisions of critical sub-phase (b)

Interestingly, critical sub-phase (b) can itself be further subdivided into two regions of distinct behaviour. These can be classified by regarding the number of highly occupied sites that the weak peak in $\mathrm{p}(n)$ represents and what contribution these sites give to the overall number of particles in the system. These sites will have average occupations of around $n^\star \sim L^{k/(k+1)}$ and the number of them is estimated by simply multiplying the weight of the weak peak by L. Thus the expected number of such sites goes as $L^{1-s+k/(k+1)}$. Now the average number of particles in the system is given by

$$\langle N \rangle = L \int \mathrm{d}n\, n\, \mathrm{p}(n)\,, \tag{5.60}$$

so the contribution from the critical, power-law part of the distribution is

$$\langle N \rangle_{\text{crit}} \sim L \int \mathrm{d}n\, n^{1-b} \sim \mathcal{O}(L)\,, \tag{5.61}$$

and the contribution from the weak peak is

$$\langle N \rangle_{\text{wkpk}} \sim Ln^{\star} \times \text{weight} \sim L^{2k/(k+1)+1-s} \,. \tag{5.62}$$

Thus for $s < 2k/(k+1)$ the weak peak gives the dominant contribution to the number of particles in the system and while being a critical phase in terms of the p(n) analysis it actually has a density that is greater than the critical density—this will be referred to as critical sub-phase (b) (ii). For $s > 2k/(k+1)$ the critical, power-law part of p(n) gives the dominant contribution to the number of particles in the system and the density will tend to the critical density asymptotically—this will be referred to as critical sub-phase (b) (i). It is also interesting to note that within the allowed values of s the number of sites represented by the weak peak is less than one for $kb/(k+1) > s > (2k+1)/(k+1)$ and when $b > 2 + 1/k$. This signals a system that may oscillate between having and not having an identifiable 'meso-condensate', i. e. a site that has a large number of particles but cannot be counted as a true condensate.

5.2.5 Generalisations

The dynamics (5.1,5.2,5.3) were chosen to show the existence of the critical phase, whilst having a particularly simple form and leaving the analysis relatively straightforward. The type of behaviour shown, i. e. the critical phase, is not restricted to these precise forms of the dynamics. In fact, exactly the same phase behaviour will be found for any event rates that have the same asymptotics as (5.1,5.2,5.3). This includes any power-series for the rates that have (5.1,5.2,5.3) as their leading terms in L. It is also straightforward to generalise to creation rates that depend on the number of particles at a site. This complicates the analysis and can introduce new behaviour, but the core element of the critical phase can remain for some choices for the dependence on particle number.

It should be noted that in order to have the power-law in the conserving system, the hop-rate must have the same asymptotics as (5.1). However, when the non-conservation is introduced, the annihilation rate enters into the expression for p(n) (5.11) in much the same way as the hop rate. Thus it is entirely plausible that the power-law could emanate from the annihilation instead of the hopping. Although not entering the analysis in the same way, it is equally plausible that the power-law could emanate from a creation rate that depends on the number of particles at a site in an appropriate way.

Coming back to the case where the hop rate provides the power-law behaviour, the main ingredients then required to generate a critical phase are as follows. Either a creation rate that increases at lowly occupied sites, or an annihilation rate that

increases at highly occupied sites, or a mixture of both. These rates must also have system size dependence that any bumps or cut-offs they introduce happen at a position that scales with the system size. Under these conditions there may be other choices of hop rate that exist that can cause the system to have a critical phase rather than just a critical point.

5.2.6 Numerical simulation results

In this section results from simulation of the model are presented and compared with the theory. The dynamics of this system (5.1), (5.2) and (5.3) is conveniently implemented by a simple Monte Carlo algorithm. At each time step, separated by Δt, the following update processes take place:

- A site is selected at random.
- With probability $u(n)\Delta t$, where $u(n)$ is given by (5.1), a particle is removed and replaced at another randomly chosen site. This is the hopping process.
- With probability $a(n)\Delta t$, where $a(n)$ is given by (5.2), a particle is removed from the site. This is the annihilation process.
- With probability $c\,\Delta t$ a particle will be added to the site, where c is given by (5.3). This is the creation process.

The time step size, Δt, is a constant and is chosen such that the probability of any one of these processes happening in a given time step, i. e. $(u(n)+a(n)+c)\Delta t$, is less than or equal to one. The code for the simulation is presented in Appendix E. Note that the annihilation rate (5.2) is difficult to normalise as it can become infinite. For simulations this rate is capped at its value at $n = L$. For all the data presented, a site with L or more particles was never realised over the duration of the simulation and so the capping of the annihilation rate had no effect.

All results presented in this section are taken from simulations run for $\mathcal{O}(10^7)$ Monte Carlo sweeps, where a sweep is L time steps. In all cases the first half of the run was used to relax the system to the steady state and the second half was used to calculate the occupancy distribution, p(n). Simulations were run on a lattice of 5000 sites with $b = 2.6$ and $k = 3$ and the following s values to demonstrate the different phases.

- critical sub-phase (a): $s = 2$.
- critical sub-phase (b) (i) : $s = 1.7$.

- critical sub-phase (b) (ii) : $s = 1.2$.
- high density phase: $s = 0.4$.

The occupancy distributions for these parameters are given in Figure 5.2 and compared with theory lines from (5.33), (5.41) and (5.59). No data are presented for the low density phase as the steady-state relaxation timescale proved prohibitive for the L, b, k parameter values chosen here. Unfortunately it is difficult to find parameters that give a low density phase with an attainable relaxation that do not also have strong finite-size effects affecting some of the other phases. Such data are presented in Section 5.3.7, Figure 5.7 where the finite-size effects will be discussed in more detail.

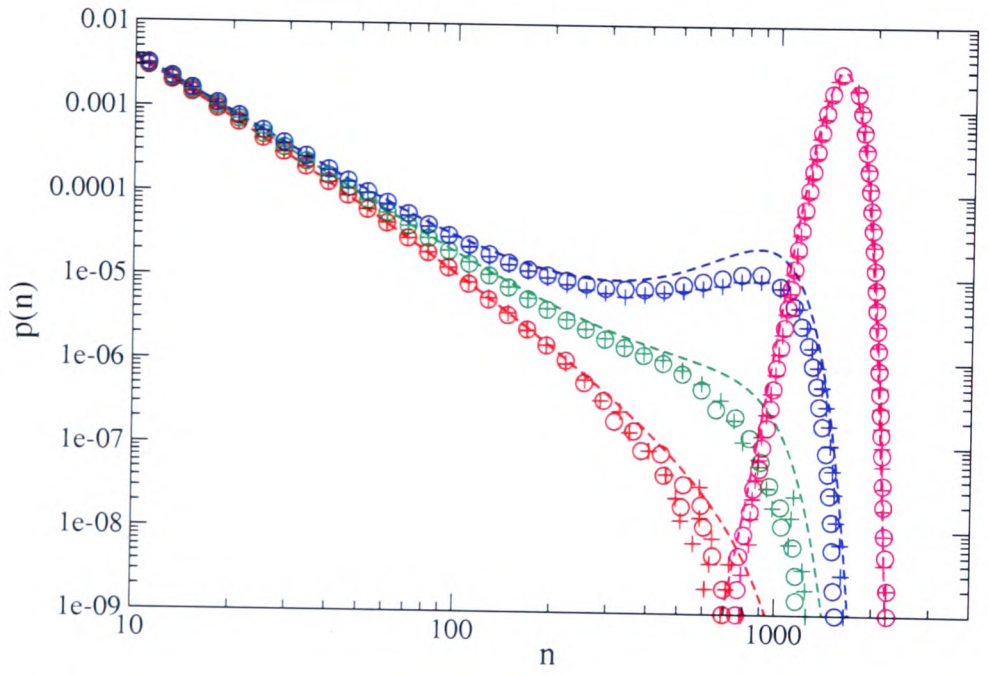


Figure 5.2: Steady state distributions of the number of particles at a site from simulations of the non-conserving ZRP model on a fully connected lattice (◯) and a 1d lattice (+), compared with theoretical asymptotic curves (dashed lines). Simulations were run of a system with $L = 5000$ lattice sites and $b = 2.6$, $k = 3$. Data are shown for: critical sub-phase (a), $s = 2$ (red); critical sub-phase (b) (i), $s = 1.7$ (green); critical sub-phase (b) (ii), $s = 1.2$ (blue); high density phase, $s = 0.4$ (magenta).

In Figure 5.2 data are also presented for a system defined on a one-dimensional lattice where particles can only hop to the right neighbour site in. While the steady state probability distribution is not expected to factorise, even in the infinite system size limit, the data for the two systems compare well. The data from simulations also compares very well with the theory lines. There is some discrepancy for critical sub-phase (b) (i) and (ii), but this is probably due to the fact that certain logarithmic corrections have been ignored and the finite size of the system. It is expected that

the agreement between theory and simulation will get better as the system size is increased.

5.3 Application to networks

5.3.1 Introduction

The general applicability of the ZRP as a network model has been discussed in Chapter 4, Section 4.7 and the generalised non-conserving ZRP introduced above is no different in its suitability to describe a network model. There are, however, subtle differences that are introduced when the model is applied to networks. In the mapping, sites are again interpreted as nodes and particles are again interpreted as the ends of links. However, it is no longer possible to create and annihilate single particles and maintain a proper network structure, instead pairs of particles must be created and annihilated—this corresponds to the creation and annihilation of links—and the pairing of the particles must be tracked throughout the evolution of the system.

Thus the following network model is considered. There are L nodes, the number of which is conserved, that are linked together by $N/2$ undirected links, the number of which is not conserved and with N being the number of particles in the corresponding ZRP. Links will rewire away from a node to another randomly selected node with rate $u(n_\mu)$, which depends on the degree of the node that the link is being rewired from, n_μ. Only the end of the link at node μ changes in the rewiring move; the other end remains attached to the same node. Links will be annihilated from a node with rate $a(n_\mu)$, which also depends on the degree of the node that the link is being annihilated from, n_μ. Links are obviously annihilated at both ends, but the rates are defined such that links are annihilated depending on one end of the link. Finally links will be created between two randomly chosen nodes with a rate c, which is a constant. These rules have been chosen to maintain a strong link with the ZRP model. The subtle differences that come from working in the network context will be discussed below. For simplicity the following explicit rates are studied.

$$u(n) = \left(1 + \frac{b}{n}\right)\theta(n)\,, \tag{5.63}$$

$$a(n) = \left(\frac{n}{L}\right)^k\,, \tag{5.64}$$

$$c = \frac{1}{L^s}\,. \tag{5.65}$$

In keeping with the desire to maintain a strong link between the two systems the rates are the same as for the ZRP model.

5.3.2 Analysis

As the network model has also been defined on a fully-connected geometry, i. e. a link can rewire to any node with equal probability, it is also expected that the steady state degree distribution of this system will factorise as before, $\mathrm{P}(n_1, n_2, \ldots, n_L) \propto \prod_{i=1}^{L} \mathrm{p}(n_i)$. Thus a mean-field master equation in the degree of a site is written for the system in the same way as for the ZRP model, but due to the creation and annihilation of pairs of particles it differs slightly and is given by

$$\begin{aligned}\frac{\partial \mathrm{p}(n)}{\partial t} = &[u(n+1) + a(n+1) + \Lambda(n+1)]\mathrm{p}(n+1) - [v+2c]\mathrm{p}(n) \\ &- \{[u(n) + a(n) + \Lambda(n)]\mathrm{p}(n) - [v+2c]\mathrm{p}(n-1)\}\,\theta(n)\end{aligned} \tag{5.66}$$

The square bracketed terms of the above equation, coupled with the probabilities, have the same interpretation as for the ZRP master equation (5.4): the first term represents the probability current into a state with degree n due to rewirings or annihilations from nodes with degree $n+1$; the second term represents the probability current out of a state with degree n due to rewirings from and creations at nodes with degree n; the third term represents the probability current out of a state with degree n due to rewirings or annihilations from nodes with degree n; and the final term represents the probability current into a state with degree n due to rewirings into or creations at nodes with degree $n-1$.

There are two differences between the network master equation (5.66) and the ZRP master equation (5.4): c has become $2c$ in the network master equation, this is because particles are created in pairs so there is twice the chance of a degree-increasing creation event occurring; the other difference is the inclusion of a new term $\Lambda(n)$ that accounts for the fact that links are annihilated rather than individual particles. The annihilation part of the dynamics is defined such that links are annihilated from nodes with a rate that depends on the degree of that node. Thus when a link is annihilated it will enter the master equation as $a(n)$ for the node that it was annihilated from. The annihilation must also enter the master equation for its other end and even within an uncorrelated, mean-field framework the degree of the node at the other end of the link will enter the equation. This is accounted for by the term $\Lambda(n)$, which is given by

$$\Lambda(n) = \frac{n}{N}\sum_{l=1}^{\infty} a(l)\mathrm{p}(l) = c\frac{n}{N}\,. \tag{5.67}$$

In the mean-field framework, a link chosen at one end will be attached to a node with degree n at the other end with probability equal to $n/(N-1) \sim n/N$. Thus the term $\Lambda(n)$ above is given by the probability that a chosen link has degree n at its other end

multiplied by the overall average rate of annihilation of links at the chosen ends. In this respect $\Lambda(n)$ is the average annihilation at sites of degree n due to the non-chosen ends of the links and is similar in nature to the term v which represents the average current due to rewirings away from other sites. The average current v takes the same form as for the ZRP version

$$v = \sum_{n=1}^{\infty} u(n)\mathrm{p}(n) \,, \tag{5.68}$$

as do the normalisation and creation-annihilation balance conditions

$$\sum_{n=0}^{\infty} \mathrm{p}(n) = 1 \tag{5.69}$$

$$\sum_{n=1}^{\infty} a(n)\mathrm{p}(n) = c \,. \tag{5.70}$$

Note that the $\mathrm{p}(n)$ dependency is present for v, but can be summed out for $\Lambda(n)$ using the creation and annihilation balance condition (5.70), as done in (5.67).

The steady state degree distribution $\mathrm{p}(n)$ can be found in a similar way to before (5.6)-(5.11), to give

$$\mathrm{p}(n) = \frac{(v+2c)^n}{\prod_{m=1}^{n}[u(m)+a(m)+\Lambda(m)]}\,\mathrm{p}(0) \,. \tag{5.71}$$

The differences between the two models actually make no difference in the asymptotic analysis of the phases. The constant c in the numerator of (5.71) only entered asymptotically in the ZRP analysis and a factor of two in front makes no difference. The new term $\Lambda(n)$ merely induces an extra cut-off, which happens never to dominate over the others. To show this, it is sufficient to assume the analysis from the ZRP model is valid and check that the new-cut off does not dominate consistently with the analysis. Thus the regions of behaviour for the network model are the same as for the non-conserving ZRP model and this is shown in the following section.

5.3.3 Phase analysis for the non-conserving network model

The network model creation-annihilation balance equation can be written as follows

$$L^{k-s} = \sum_{n=1}^{\infty} n^k \mathrm{p}(n) \tag{5.72}$$

$$= \mathrm{p}(0) \sum_{n=1}^{\infty} n^k \exp\left[n\ln(v+2c) - \sum_{m=1}^{n} \ln(u(m)+a(m)+\Lambda(m)) \right] \,. \tag{5.73}$$

Hence the phase analysis is again easily split into two regions depending on whether the left hand side of the above equation converges or diverges.

Low density phase: $s > k$

The analysis for the low density phase in the non-conserving ZRP model was independent of all quantities except the current v and the new term $\Lambda(m)$ introduced in the network model makes no difference to this. Thus the low density phase still has an exponentially decaying $\mathrm{p}(n)$, in exactly the same way as for the non-conserving ZRP, which is found to be approximated well by

$$\mathrm{p}(n) \sim \exp\left[n \ln(L^{k-s})\right] , \tag{5.74}$$

for large but finite systems.

Analysis of region: $s < k$

In this region, as with the ZRP case, the creation-annihilation balance equation (5.73) is dominated by terms at high n in the sum and $\sum \ln(u(m) + a(m) + \Lambda(m))$ is approximated as

$$\begin{aligned} \sum_{m=1}^{n} \ln(u(m) + a(m) + \Lambda(m)) &\sim \int_1^n \mathrm{d}m \left[\frac{b}{m} + \left(\frac{m}{L}\right)^k + \frac{m}{NL^s}\right] \\ &\sim b \ln n + \frac{n^{k+1}}{(k+1)L^k} + \frac{n^2}{2NL^s} . \end{aligned} \tag{5.75}$$

Thus the balance equation (5.73) can be written as

$$L^{k-s} \sim \int_1^\infty \mathrm{d}n\, n^{k-b} \exp\left[n g(L) - \frac{n^{k+1}}{(k+1)L^k} - \frac{n^2}{2NL^s}\right] , \tag{5.76}$$

where $v + 2c \sim 1 + g(L)$ has been assumed. This is the same as the equivalent equation for the ZRP case, with the exception of the extra cut-off. In the following it is shown that this new cut-off is never dominant and the asymptotic analysis of the ZRP model is sufficient to describe the network model as well.

High density phase: $s < k/(k+1)$

In the high density phase of the non-conserving ZRP, $\mathrm{p}(n)$ was found to behave asymptotically as (5.33)

$$\mathrm{p}(n) \sim n^{-b} \exp\left[g_h \frac{n}{L^s} - \frac{n^{k+1}}{(k+1)L^k}\right] , \tag{5.77}$$

and that the average number of particles in the system behaved as $\langle N \rangle \sim L^{2-s/k}$. Thus if the cut-off given by $\exp(-n^{k+1}/((k+1)L^k))$ dominates over the new cut-off, given by $\exp(-n^2/(2NL^s))$, then the ZRP analysis remains correct for the network model. This will be so when

$$L^{k/(k+1)} < N^{1/2} L^{s/2} = L^{1+s(k-1)/2k} , \tag{5.78}$$

i. e. when

$$\frac{k}{k+1} < 1 + \frac{s(k-1)}{2k} \,, \tag{5.79}$$

which is always true in the region $s < k/(k+1)$. Thus the high density phase of the network model is the same analytically as for the ZRP model; in the networks context, the degree distribution is strongly peaked around $n \sim L^{1-s/k}$ and there is a super-extensive average number of links in the system $\langle N \rangle /2 \sim L^{2-s/k}$.

Critical sub-phase (a): $k > s > kb/(k+1)$

In critical sub-phase (a) of the non-conserving ZRP, $\mathrm{p}(n)$ was found to behave asymptotically as

$$\mathrm{p}(n) \sim n^{-b} \exp\left[-g_{ca} \frac{n}{L^{(k-s)/(k-b+1)}}\right] \,. \tag{5.80}$$

The average number of link ends in this phase is calculated to be

$$\begin{aligned} \langle N \rangle &\sim L \int_1^\infty \mathrm{d}n \, n\mathrm{p}(n) \\ &\sim L \int_1^{L^{(k-s)/(k-b+1)}} \mathrm{d}n \, n^{1-b} \\ &\sim L \,, \end{aligned} \tag{5.81}$$

where it has been assumed that $b > 2$, as in the ZRP case. The cut-off, given by $\exp(-g_{ca} n/L^{(k-s)/(k-b+1)})$, dominates over the new cut-off, given by $\exp(-n/2NL^s)$, when

$$L^{(k-s)/(k-b+1)} < N^{1/2} L^{s/2} \sim L^{(1+s)/2} \,, \tag{5.82}$$

i. e. when

$$k + b - 1 < s(k - b + 3) \,, \tag{5.83}$$

which is always true in the region $k > s > kb/(k+1)$ and with $k > b-1$. Thus critical sub-phase (a) of the network model is the same, analytically, as for the ZRP model. The degree distribution is a power-law with an exponential cut-off and the number of links goes like L. In the thermodynamic limit the density of links will tend to the half of the critical density of particles in the corresponding conserving ZRP.

Critical sub-phase (b): $kb/(k+1) > s > k/(k+1)$

In critical sub-phase (b) of the non-conserving ZRP, $\mathrm{p}(n)$ was found to behave asymptotically as

$$\mathrm{p}(n) \sim n^{-b} \exp\left[n \left\{ \left(b - \frac{s(k+1)}{k} \right) \frac{\ln L}{L} \right\}^{k/(k+1)} - \frac{n^{k+1}}{(k+1)L^k} \right] \,. \tag{5.84}$$

In this sub-phase the number of links in the system can have two asymptotic behaviours: in the region $kb/(k+1) > s > 2k/(k+1)$ the power-law part of the degree distribution, p(n), dominates the number of links, which behaves as L; in the region $2k/(k+1) > s > k/(k+1)$, the weak peak part of the degree distribution dominates the number of links which behaves as $L^{2k/(k+1)+1-s}$. Clearly the smallest that $\langle N \rangle$ can be is $\mathcal{O}(L)$ and if the cut-off, $\exp(-n^{k+1}/((k+1)L^k))$, dominates over the new cut-off, $\exp(-n^2/2NL^s)$, with the smallest possible value of N throughout the sub-phase, then the ZRP analysis will be proven to be valid throughout this phase. The condition for this is

$$L^{k/(k+1)} < N^{1/2} L^{s/2} \sim L^{(1+s)/2} , \tag{5.85}$$

i. e.

$$s > \frac{k}{k+1} - \frac{1}{k+1} , \tag{5.86}$$

which is always true in the region $kb/(k+1) > s > k/(k+1)$. Thus critical sub-phase (b) of the network model is the same as for the ZRP model, analytically. The degree distribution is a power-law with a weak peak at high n which does not dominate the distribution, but does dominate the annihilation, and for $2k/(k+1) > s > k/(k+1)$ it also dominates the average number of links in the system. Thus in the region $kb/(k+1) > s > 2k/(k+1)$ the density of links in the system relaxes to half the critical density of particles in the corresponding ZRP; in the region $2k/(k+1) > s > k/(k+1)$ the density of links is greater than the critical density, but the phase is still critical from the mathematical viewpoint of the phase analysis.

5.3.4 Numerical simulation results

In the ZRP model, only knowledge of the number of particles at each site is needed to give a complete description of the system. When simulating the network model the network structure itself must be considered in order to correctly account for the deletion of links: just deleting two particles at random is not enough. One method of keeping track of a network computationally, that is favourable to simulate the network model described above, is as follows.

- Two arrays are used each holding pointers to information for each link.
- Each link is stored with information on which nodes it is connected to and where the pointer to the link is stored in the arrays.
- One of the arrays is one-dimensional and holds pointers to the links; this can be used to choose a link at random, if this is desired.

- The other array is two-dimensional and the i, j element holds a pointer to the j^{th} link that is attached to node i. Each link is therefore held twice and not all elements of the array hold pointers to actual links, some hold null pointers. This array holds a complete description of the network structure which allows the links to be deleted correctly, i. e. to find the other end of a link which has been chosen by one of its ends at a node.

The fact that a complete description of the network is stored in principle allows an output of the network state at any point in the evolution and calculation of properties other than the degree distribution. However, such features were not implemented. The calculation of specific network properties is often most efficiently achieved by storing values of the parameter of interest and updating as the network evolves rather than updating the network and calculating the values of the parameter at given time intervals. In this case the degree distribution is the quantity of interest and is held in a separate array and averaged over.

The original code for the network model was supplied by Erel Levine and is listed in Appendix F, in the original version self-links were not allowed as this makes the code somewhat simpler. However, the disallowal of self-links and multiple links can have a noticeable effect on the system, see Section 5.3.5. A version of the code modified to allow for self-links is given in Appendix G

The core of the program is a simple Monte Carlo algorithm, very similar to that in the simulations of the non-conserving ZRP above. At each time step, separated by Δt, the following update processes take place:

- A node is selected at random and if this node has any links one is chosen at random. This corresponds to randomly choosing a row of the two-dimensional array and then choosing one of the elements of this row that contains a pointer to a link at random.

- With probability $u(n)\Delta t$, where $u(n)$ is given by (5.63), this link is rewired away from this node to another randomly chosen node. This is the link rewiring process. This requires entries in all of the arrays to be suitably updated.

- With probability $a(n)\Delta t$, where $a(n)$ is given by (5.64), this link is removed from the network. This is the link annihilation process. Again this requires suitable updating of all of the arrays.

- With probability $c\Delta t$, where c is given by (5.65), a link connected to this node and another randomly selected node is added to the network. This is the link

creation process. This requires new entries to all of the arrays to account for the new link.

Again the time step size, Δt is constant and is chosen such that the probability of any one of these events happening in a single time step is less than or equal to one. The majority of the complexity of the program is contained in correctly updating the data structures that store the network-specific information.

The results presented in this section are taken from simulations run for $\mathcal{O}(10^7)$ Monte Carlo sweeps. To enable direct comparison with the non-conserving ZRP model the same system size and parameters were used, namely a network of $L = 5000$ nodes with $b = 2.6$ and $k = 3$ and the s values: $s = 2$, to show critical sub-phase (a); $s = 1.7$, to show critical sub-phase (b) (i); $s = 1.2$ to show critical sub-phase (b) (ii); and $s = 0.4$ to show the high density phase. The results from the simulations are presented in Figure 5.3, with theory lines for comparison. The network and non-conserving ZRP data are both plotted in Figure 5.4, this allows a more direct comparison of the two systems. Again no data are presented for the low-density phase as relaxation time-scales are prohibitive for these values of b, k and L and the values of s that they allow for the low-density phase. Parameter values that give accessible relaxation times for the low-density phase are possible, but they give their own complications for some of the other phases, as discussed in Section 5.3.5.

It can be seen from Figure 5.3 that the network model degree distributions, $\mathrm{p}(n)$, compare reasonably well with the theory lines as calculated within the ZRP framework and expected analytically to also describe the network model. However there are some noticeable discrepancies for the high density phase and critical sub-phase (b) (ii). A possible explanation for this is the effect of degree-degree correlations in the system, where the likelihood that two nodes are connected depends on the degree of each node in a non-trivial fashion, which are not treated by the mean-field analysis. Degree-degree correlations are known to be present in many real networks and models—see for example [57]. One instance of this is in the Internet where it is found that high degree nodes are more likely to be connected to nodes with low degree [54, 55]. It could be that the rewiring and annihilation dynamics of this model are somehow generating degree-degree correlations that are not accounted for in the theory.

However, despite these discrepancies the phases appear to be qualitatively the same: curves linking the data points are the same shape, i. e. the power-law with a weak peak and strongly peaked forms for $\mathrm{p}(n)$ are retained.

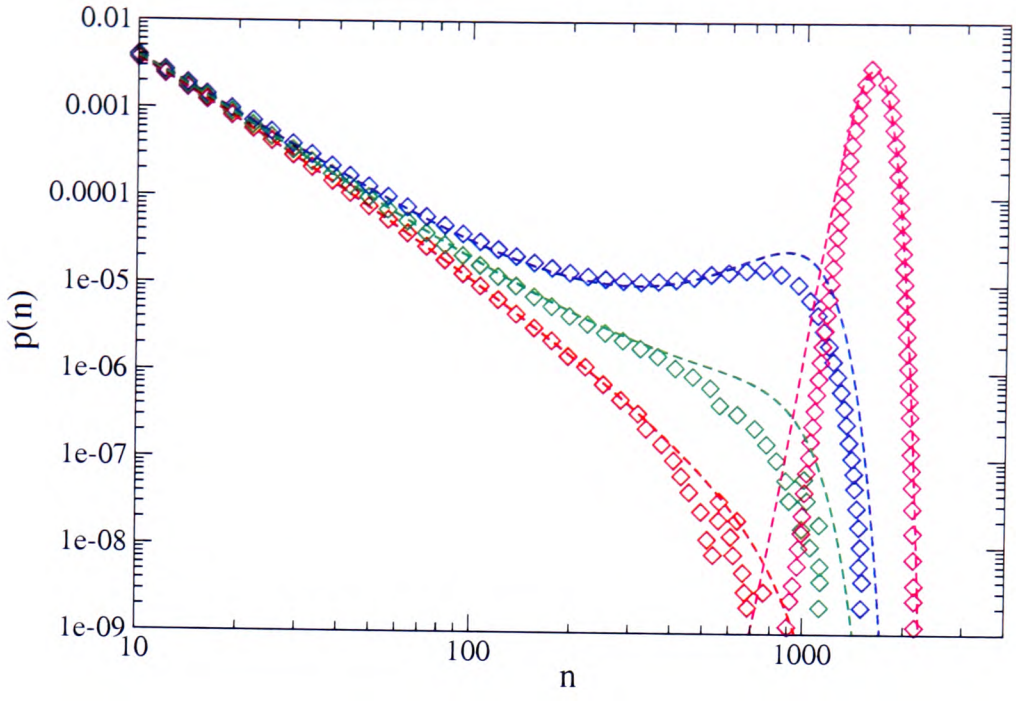


Figure 5.3: Steady state degree distributions from simulations of the non-conserving network model with self-links allowed (◊) compared with theoretical asymptotic curves (dashed lines). Simulations were run on a network with $L = 5000$ nodes with $b = 2.6$, $k = 3$. Data are shown for: critical sub-phase (a), $s = 2$ (red); critical sub-phase (b) (i), $s = 1.7$, (green); critical sub-phase (b) (ii), $s = 1.2$ (blue); high density phase, $s = 0.4$ (magenta).

5.3.5 Effect of disallowing self-links and multiple links

In the networks model the analysis has assumed that self links and multiple links are present. The absence of these is known to generate degree-degree correlations under certain circumstances. For example, consider the equilibrium rewiring network of Dorogovtsev, Mendes and Samukhin [139, 8]. This model has strong connections with the basic ZRP as a network model and can be studied within the kind of mean-field approximation (which happens to be exact in the conserving case) as in this chapter [8]. To reconstruct the network in this picture the average can be taken over all the possible networks constructed from a given degree sequence or distribution. This becomes considerably more involved when self links and multiple links are prohibited. In the uncondensed case, the number of self and multiple links generated, i. e. the number of configurations with such links, is sufficiently small that it makes little difference to the ensemble average. In the condensed case, where a single node captures a finite fraction of the available link-ends, the fraction of constructed network states that contain self and multiple links becomes appreciable and thus the analysis with or without such links differs considerably. For example, if the condensed node contains more than half of the available link ends, then it is not possible to construct a network state without

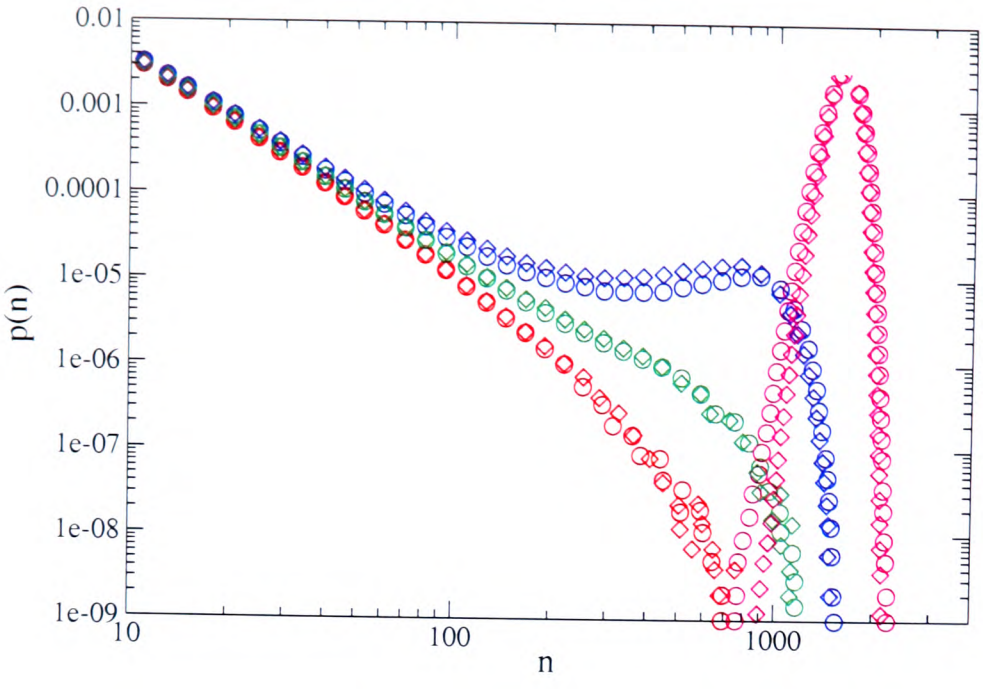


Figure 5.4: Steady state distributions of the number of particles at a site and degree from simulations of the non-conserving ZRP (◯) and the non-conserving network model with self-links allowed (◇). Simulations were run on systems with $L = 5000$ sites or nodes with $b = 2.6$ and $k = 3$. Data are shown for: critical sub-phase (a), $s = 2$ (red); critical sub-phase (b) (i), $s = 1.7$, (green); critical sub-phase (b) (ii), $s = 1.2$ (blue); high density phase, $s = 0.4$ (magenta).

multiple self links at the condensed site. Thus with this constriction, correlations in the system become too strong and the mean-field analysis is no longer valid. The question of random networks without self and multiple connections has recently been addressed rigorously in [144]. The absence of self and multiple connections, in terms of random uncorrelated networks, has also been discussed in [145]. Correlations in exponential random graph models that have given degree sequences and multiple links but no self links have been discussed in [135]. The absence of multiple links has also been seen to partially explain the degree-degree correlations observed in the Internet in [56, 146].

In some original simulations of the network model, self-links were disallowed for computational reasons. It was found that doing so strongly affected the results. The original simulations redirected any rewiring or creation event that created a self-link, to another node until no self link was created. This mechanism can be partially accounted for within the mean-field master equation approach.

For link creation, the disallowal of self links makes no difference to the master equation; a creation adds two link ends and the probability that the two ends are placed at the same node depends only on the number of the nodes, thus the redirection is shared equally among all nodes. For link rewiring however, the probability that the

rewired end of the link is rewired to the node to which its other end is attached does depend on the degree of that node and hence the configuration of the system. Thus the disallowal of self links enters the master equation through the rewiring rate. and it is Note also that an annihilation event cannot create a self link. Remaining within the uncorrelated mean-field approximation, the probability that a link chosen randomly at one end has its other end at a node of degree n is proportional to n/N. Hence the probability of a redirection is given by $(1-n/N)/\langle 1-n/N\rangle = (1-n/N)/(1-1/L)$. This effect is then incorporated into the master equation by replacing the incoming current due to hops, v, with a new n-dependent version, $v(n)$, that accounts for the effect of redirections. The iterative solution of the master equation (5.71) then becomes

$$\mathrm{p}(n) = \frac{\prod_{m=1}^{n}[v(m)+2c]}{\prod_{m=1}^{n}[u(m)+a(m)+\Lambda(m)]}\,\mathrm{p}(0)\ , \tag{5.87}$$

where $v(m)$ is given by

$$v(m) = v\frac{1-m/N}{1-1/L} = \frac{1-m/N}{1-1/L}\sum_{n=1}^{\infty}u(n)\mathrm{p}(n)\ . \tag{5.88}$$

The phase analysis then proceeds from the creation annihilation balance equation (5.73) which now takes the form

$$L^{k-s} = \mathrm{p}(0)\sum_{n=1}^{\infty}n^{k}\exp\left[\sum_{m=1}^{n}\{\ln(v(m)+2c)-\ln(u(m)+a(m)+\Lambda(m))\}\right]\ . \tag{5.89}$$

The new addition to $\mathrm{p}(n)$ due to the disallowal of self links substantially affects the analysis. The three main phases remain the same; however, the critical phase has a new catalogue of subdivisions. The following behaviour is found

- **Low density phase, no self links** $(s > k)$: In this region the degree distribution is exponentially decaying and the global density of particles tends to zero in the thermodynamic limit.

- **Critical phase, no self links** $(k/(k+1) < s < k)$: In this region the probability has a power-law form in the thermodynamic limit. The large but finite behaviour of the degree distribution can be used to classify two sub-phases and three regions of the second sub-phase.

  - **Critical sub-phase (a), no self links** $(k > s > (k+b-1)/2)$: In this sub-phase the degree distribution is a power law with an exponential cut-off.

  - **Critical sub-phase (b), no self links** $((k+b-1)/2 > s > k/(k+1))$: In this sub-phase the degree distribution is a power law with a weak peak

at high n. The form of the weak peak can be used to further classify three subdivisions of this region.

  * **Subdivision (i)** $((k+b-1)/2 > s > (k+1)/2)$: In this subdivision the shape of the weak peak is due to the new cut-off and the number of link ends in the system scales as the system size, i. e. the weak peak does not dominate the number of links in the system.
  * **Subdivision (ii)** $((k+1)/2 > s > 1)$: In this subdivision the shape of the weak peak is due to the new cut-off and the number of link ends in the system scales super-linearly with the system size, i. e. the weak peak dominates the number of links in the system.
  * **Subdivision (iii)** $(\min(1,k) > s > k/(k+1))$: In this subdivision the shape of the weak peak is due to the old cut-off and is the same in form as for critical sub-phase (b) (ii) of the ZRP and network with self links allowed models.

- **High density phase, no self links** $(s < k/(k+1))$: In this region the degree distribution tends to a delta peak and all the nodes have high degree.

The phase diagram for this system is presented in Figure 5.5. The analysis that identifies these phases and subdivisions is discussed in detail in the following.

Low density phase, no self links $(s > k)$

Again, in the low density phase the decay of $\mathrm{p}(n)$ is so rapid, to cope with the LHS of (5.89) going to zero, that the detailed form of (5.87) doesn't really play a role in the analysis. Thus the degree distribution is the same as that from the analysis of the network with self links and ZRP models

$$\mathrm{p}(n) \sim \exp\left[n \ln(L^{k-s})\right] . \tag{5.90}$$

Thus this phase is the same as for the network and ZRP models. For the purpose of distinguishing the no self links phases from the with network model with self links in the cases where they do differ, this phase will be referred to as the NSL (no self links) low density phase.

Analysis of region $(s < k)$

For $s < k$ the LHS of (5.89) diverges and so the expression for $\mathrm{p}(n)$ (5.87) is expanded by approximating the sums as integrals and assuming that $v - 1 \sim g(L)$ as before,

giving

$$\mathrm{p}(n) \sim n^{-b} \exp\left[ng(L) - \frac{n^2}{2N} - \frac{n^2}{2N}g(L) + \frac{2n}{L^s} - \frac{2n}{L^{s+1}} - \frac{n^2}{2NL^s} - \frac{n^{k+1}}{(k+1)L^k} + \frac{n}{L}\right] , \quad (5.91)$$

where the new addition has caused many extra factors to appear in this expression. The most significant of these is $\exp(-n^2/2N)$. So an effective form for $\mathrm{p}(n)$ can be written

$$\mathrm{p}(n) \sim n^{-b} \exp\left[ng(L) - \frac{n^2}{2N} - \frac{n^{k+1}}{(k+1)L^k}\right] , \quad (5.92)$$

and the question becomes: which of these cut-offs will dominate? Unlike the cut-offs introduced by $\Lambda(n)$, this new cut-off can become dominant. All the phases remain qualitatively in the presence of the new cut-off, i. e. the power-law with a cut-off, power-law with a weak peak and the strongly peaked forms for $\mathrm{p}(n)$ are recovered. However, the cut-offs and peaks do not necessarily have exactly the same form as the network model with self links allowed and some of the phase and sub-phase boundaries are shifted.

Critical sub-phase (a), no self links $(k > s > (k + b - 1)/2)$

In this sub-phase $g(L)$ has the form

$$g(L) = -\frac{g_{ca}}{L^x} , \quad (5.93)$$

where g_{ca} is a positive constant, and this generates the dominant cut-off, giving the balance equation

$$L^{k-s} \sim \int_1^\infty \mathrm{d}n\, n^{k-b} \exp\left[-g_{ca}\frac{n}{L^x}\right] \sim L^{x(k-b+1)} . \quad (5.94)$$

To satisfy this balance equation requires $x = (k - s)/(k - b + 1)$. This phase has the same form for $\mathrm{p}(n)$ as the network with self links and ZRP versions of the model,

$$\mathrm{p}(n) \sim n^{-b} \exp\left[-g_{ca}\frac{n}{L^{(k-s)/(k-b+1)}}\right] , \quad (5.95)$$

and the average number of link ends has already been calculated to be

$$\langle N \rangle \sim L . \quad (5.96)$$

Thus the assumed cut-off dominates over the new cut-off only if the following inequality is satisfied,

$$L^{(k-s)/(k-b+1)} < N^{1/2} \sim L^{1/2} , \quad (5.97)$$

i. e. when

$$s > \frac{(k+b-1)}{2} \,. \tag{5.98}$$

Hence the new cut-off has altered the phase boundary. Nevertheless, this phase has exactly the same behaviour as the critical sub-phase (a) for the previous versions of the model, it just has a different phase boundary. To differentiate, this phase will be referred to as NSL critical sub-phase (a).

Critical sub-phase (b) (i), no self links ($(k+b-1)/2 > s > (k+1)/2$)

In this subdivision of sub-phase (b), $g(L)$ has the form

$$g(L) = \left(\frac{g_{cbi} \ln L}{L}\right)^{1/2} \,, \tag{5.99}$$

where g_{cbi} is a positive constant, and this gives a positive term to the exponential. The dominant cut-off is the one generated by disallowal of self links, giving the balance equation,

$$L^{k-s} \sim \int_1^\infty \mathrm{d}n\, n^{k-b} \exp\left[n\left(\frac{g_{cbi} \ln L}{L}\right)^{1/2} - \frac{n^2}{2N}\right] \,. \tag{5.100}$$

Assuming the weak peak in the integrand of the RHS of the above equation dominates, the value of g_{cbi} that satisfies the balance is

$$g_{cbi} = k - 2s + b - 1 \,, \tag{5.101}$$

giving the following form for the occupancy distributions

$$\mathrm{p}(n) \sim n^{-b} \exp\left[n\left((k-2s+b-1)\frac{\ln L}{L}\right)^{1/2} - \frac{n^2}{2N}\right] \,. \tag{5.102}$$

Now, in the no self links case, critical sub-phase (b) is being subdivided by two criteria: the form of the weak peak; and the contribution of the weak peak to the number of links in the system. The subdivision of critical sub-phase (b) in this section is due to it having a weak peak with the above form (5.102) and the power-law part of the distribution dominating the number of links in the system. The average number of link ends in the system due to the critical, power-law part of the distribution is given by

$$\begin{aligned} \langle N \rangle_{\mathrm{crit}} &\sim L \int_1^{N^{1/2}} \mathrm{d}n\, n^{1-b} \\ &\sim L \,. \end{aligned} \tag{5.103}$$

for $b > 2$. The contribution from the weak peak is, as before, estimated as the weight of the weak peak multiplied by the value of the power-law at the maximum of the peak and by the number of nodes in the system.

$$\begin{aligned} \langle N \rangle_{\text{wkpk}} &\sim L(n^\star)^{1-b} \times \text{weight} \\ &\sim L^{2-b/2+g_{cbi}/2} , \end{aligned} \tag{5.104}$$

Thus the values of s for which the system is in critical sub-phase (b) i, are

$$\begin{aligned} 2 - \frac{b}{2} + \frac{g_{cbi}}{2} &< 1 \\ s &> \frac{k+1}{2} , \end{aligned} \tag{5.105}$$

to ensure that $\langle N \rangle_{\text{crit}} > \langle N \rangle_{\text{wkpk}}$ and

$$\begin{aligned} g_{cbi} &> 0 \\ s &< \frac{k+b-1}{2} , \end{aligned} \tag{5.106}$$

to ensure that there is a positive weak peak. It can easily be checked that all other conditions for the sub-phase subdivision are consistent with these values, i. e. all other inequalities that must be satisfied induce less stringent conditions on s. Thus for $(k+b-1)/2 > s > (k+1)/2$ the system is in NSL critical sub-phase (b) (i), where p(n) has the form of a power-law with a weak peak determined by the new cut-off and the system relaxes to the critical density of the underlying ZRP system.

Critical sub-phase (b) (ii), no self links ($(k+1)/2 > s > 1$)

In this subdivision of sub-phase (b), $g(L)$ has the form

$$g(L) = \frac{(g_{cbii} \ln L)^{1/2}}{L^{\theta/2}} , \tag{5.107}$$

where g_{cbii} is a positive constant to be determined. This gives a positive term in the exponential of p(n) and hence the balance equation

$$L^{k-s} \sim \int_1^\infty \mathrm{d}n\, n^{k-b} \exp\left[n \frac{(g_{cbii} \ln L)^{1/2}}{L^{\theta/2}} - \frac{n^2}{2N} \right] , \tag{5.108}$$

where again the dominant cut-off is the one generated by the disallowal of self links. Assuming that the weak peak in the integrand on the RHS of the above equation dominates, the value of g_{cbii} that satisfies the balance is

$$g_{cbii} = 2\left(b - 1 - \frac{s}{k+1} \right) , \tag{5.109}$$

giving the following form for the occupancy distribution

$$\mathrm{p}(n) \sim n^{-b} \exp\left[n \frac{(2(b-1-s/(k+1)) \ln L)^{1/2}}{L^{\theta/2}} - \frac{n^2}{2N} \right] . \tag{5.110}$$

Critical sub-phase (b) (ii) is distinguished from the other subdivisions of sub-phase (b) in that it has a weak peak that dominates the number of links in the system and with a form that is determined by the dominant new cut-off. The critical, power-law contribution to the number of link ends is calculated in the same way as before (5.103), and is given by

$$\langle N \rangle_{\mathrm{crit}} \sim L \, , \tag{5.111}$$

and the contribution from the weak peak is given by

$$\begin{aligned} \langle N \rangle_{\mathrm{wkpk}} &\sim L (n^*)^{-b} \times \text{weight} \\ &\sim L^{2-2s/(k+1)} \, . \end{aligned} \tag{5.112}$$

Thus the values of s for which the system is in critical sub-phase (b) ii, are

$$\begin{aligned} 2 - \frac{2s}{k+1} &> 1 \\ s &< \frac{k+1}{2} \, , \end{aligned} \tag{5.113}$$

to ensure that the weak peak gives the dominant contribution to the number of links in the system and

$$\begin{aligned} N^{1/2} \sim L^{1-s/(k+1)} &< L^{k/(k+1)} \\ s &> 1 \, , \end{aligned} \tag{5.114}$$

to ensure that the new cut-off is indeed the dominant cut-off. Again it is easy to check that all other required conditions are consistent with this range of possible values for s. Thus for $(k+1)/2 > s > 1$ the system is in NSL critical sub-phase (b) (ii), characterised by a $\mathrm{p}(n)$ with a weak peak that dominates both the balance equation and the number of links that are in the system. Again while this is mathematically still a part of critical sub-phase (b), the density of link ends does not relax to the critical density of particles in the underlying conserving ZRP model.

Critical sub-phase (b) (iii), no self links ($\min(1,k) > s > k/(k+1)$)

In this subdivision of sub-phase (b), $g(L)$ has the form

$$g(L) = \left(\frac{g_{cbiii} \ln L}{L} \right)^{k/(k+1)} \, , \tag{5.115}$$

where g_{cbiii} is a positive constant to be determined. The dominant cut-off is now that due to the annihilation of links from the chosen node end, i. e. the cut-off that was present even in the ZRP system, giving the balance equation

$$L^{k-s} \sim \int_1^\infty \mathrm{d}n\, n^{k-b} \exp\left[n \left(\frac{g_{cbiii} \ln L}{L} \right)^{k/(k+1)} - \frac{n^{k/(k+1)}}{(k+1)L^k} \right] . \tag{5.116}$$

Assuming the weak peak of the integrand of the RHS of the above equation dominates, the following value for g_{cbiii} satisfies the balance equation,

$$g_{cbiii} = b - \frac{s(1+k)}{k} , \tag{5.117}$$

giving the form for the occupancy distribution

$$\mathrm{p}(n) \sim n^{-b} \exp\left[n \left(\left\{ b - \frac{s(1+k)}{k} \right\} \frac{\ln L}{L} \right)^{k/(k+1)} - \frac{n^{k+1}}{(k+1)L^k} \right] . \tag{5.118}$$

Critical sub-phase (b) (iii) in the case of no self links, has the same form as critical sub-phase (b) of the ZRP and network with self links versions of the model. It is distinguished from subdivisions (i) and (ii) of NSL critical sub-phase (b) in that the original cut-off dominates the one introduced by the disallowal of self links. The contribution of the weak peak to the number of link ends is assumed to dominate and is calculated as before (5.62) to give

$$\langle N \rangle_{\mathrm{wkpk}} \sim L^{2k/(k+1)+1-s} , \tag{5.119}$$

and so the condition on the old cut-off dominating the new one is

$$\begin{aligned} L^{k/(k+1)} < N^{1/2} &\sim L^{k/(k+1)+(1-s)/2} \\ s &< 1 , \end{aligned} \tag{5.120}$$

and the condition on $g(L)$ dominating the contribution to the positive part of the exponent when compared to the contribution from $2c$ is

$$\begin{aligned} \left(\frac{L}{\ln L} \right)^{k/(k+1)} &< L^s \\ s &> \frac{k}{k+1} . \end{aligned} \tag{5.121}$$

Once more it is straightforward to check that all other required conditions are consistent with this range of possible values for s. Thus for $1 > s > k/(k+1)$ the system is in NSL critical sub-phase (b) (iii), characterised by a $\mathrm{p}(n)$ with the same form as critical sub-phase (b) of the ZRP and network with self link systems, i. e. a distribution with a weak peak that dominates the creation-annihilation balance condition and can dominate the number of link ends in the system.

Note that in the analysis of NSL critical sub-phase (b) subdivisions (i), (ii) and (iii) above a small point has been glossed over: the results given are only valid for $k > 1$. For $k < 1$ the analysis actually reverts back to that for the ZRP and network with self links variants of the model, i. e. the new cut-off does not enter significantly into the analysis.

High density phase, no self links ($k/(k+1) > s$)

As with the low density phase, the high density phase for the system without self links is analytically the same as for the ZRP or network with self links. Thus $\mathrm{p}(n)$ has the form (5.33), (5.122)

$$\mathrm{p}(n) \sim n^{-b} \exp\left[g_h \frac{n}{L^s} - \frac{n^{k+1}}{(k+1)L^k} \right] , \tag{5.122}$$

with g_h a constant that is undetermined by the analysis, but makes little difference to the asymptotic picture. Thus for $k/(k+1) > s$ the system is in a high density phase where the degree distribution is sharply peaked around $n^\star \sim L^{1-s/k}$ and the system contains a super-extensive number of link ends, $N \sim L^{2-s/k}$.

5.3.6 Discussion

The phase diagram for the system without self links is shown in Figure 5.5. It can be seen that it differs considerably from that of the system with self links allowed. It is interesting to note that the phases affected the most by the new cut-off, generated by disallowing self links, are those at or just above the critical density. This is in contrast to the non-conserving system where it is expected that problems due to self and multiple links will be seen both at the critical point and in the condensed phase, with the severity of the effect increasing with increasing density of links. This difference is due to the annihilation, which was initially introduced to evaporate any condensates in the system. In the conserving system as the density is increased the condensed node will capture a great many of the links, thus greatly increasing the chances of self and multiple links. In the non-conserving system the annihilation will prevent any one strong condensate from taking all the links, instead only weak meso-condensates will be present which increase in number as the density is increased through increasing the creation rate. These meso-condensates will tend to contain fewer self links and generally more configurations will be available that have no self links. Thus self links are less of a concern in the high density phase of this system than in the condensed phase of the basic ZRP model.

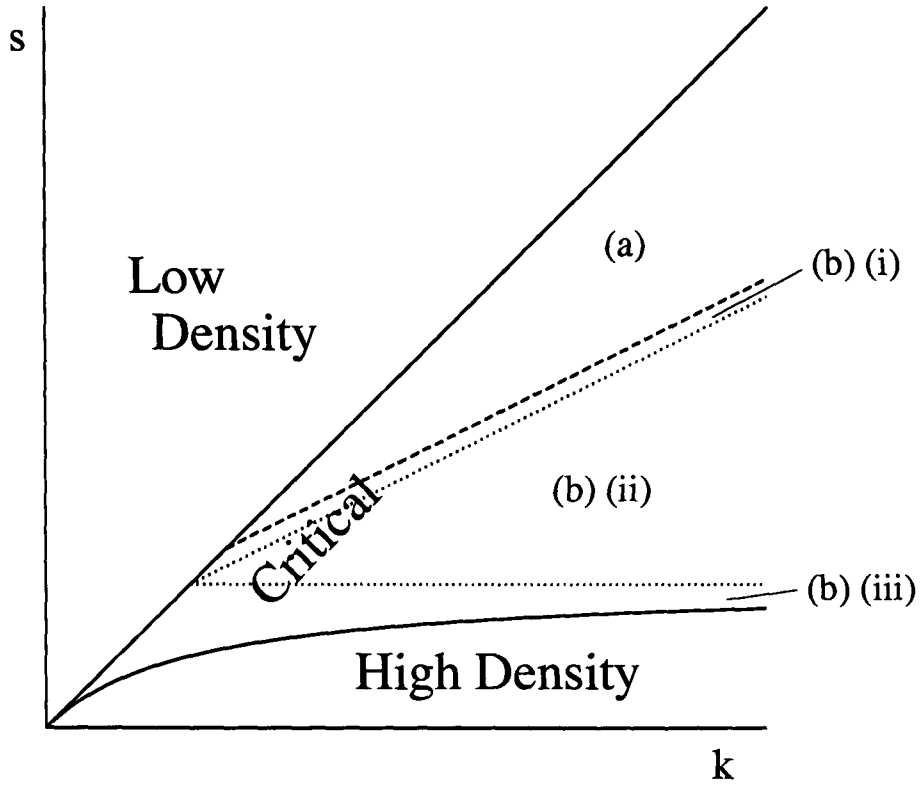


Figure 5.5: Typical phase diagram for the non-conserving network model with no self links. Shown in the $k - s$ plane with b fixed. The parameters k, s, b are defined in (5.63), (5.64), (5.65).

It is also interesting to note that, within the uncorrelated framework, the probability of a hop creating a multiple link is very similar to the probability of a hop creating a self link. Thus the analysis gives the same asymptotic results for disallowing either or both of self and multiple links. This allows the results of this section to be compared briefly with other works on cut-offs in finite-size network systems. Boguñá, Pastor-Satorras and Vespignani have discussed the effect of forbidding multiple links in general scale-free networks where the degree distribution and degree-degree correlations are known in [145]. In that work it was shown that, in order to satisfy the constraint on the maximum number of connections between nodes of degree k and k', a cut-off must be present in the degree distribution for systems of finite size. For example, the maximum number of edges between all nodes of degree k and all nodes of degree k' is $\min\{kL_k, k'L_{k'}, L_kL_{k'}\}$, where L_k is the number of nodes with degree k. The cut-off induced was found to be of the form $\exp(-n^2/2L)$, where n is the degree. The absence of self and multiple links from a statistical ensemble defined via a minifield theory was considered by Burda and Krzywicki in [126]. By considering the relative entropies of nodes of degree n with and without self and multiple links, the same cut-off was found, i. e. the factor $\exp(-n^2/2L)$ is present in the degree distribution for the system without self and multiple links when compared to that with self and multiple links. More recently Dorogovtsev, Mendes, Povolotsky and Samukhin

have rigourously calculated the effects of disallowing self and multiple links from the equilibrium rewiring model of Dorogovtsev, Mendes and Samukhin discussed in Chapter 4 Section 4.6.1 and in [144]. They find similar results and remark that the cut-offs found in [145, 126] are in fact upper estimates of the true cut-off. Specifically, they find that at the critical point of the model, i. e. where the degree distribution becomes a power-law, a cut-off $\exp(-An^2/2L)$ when $b > 3$, and a cut-off $\exp[-n^2(B/L)^{2/(5-b)}]$ when $2 < b \leq 3$, where b is the exponent of the degree distribution and A and B are constants depending on the specifics of the system. Thus the cut-off is asymptotically the same as those found previously when $b > 3$, i. e. when the degree distribution has a converging second moment.

These cut-offs can be compared with those found for the non-conserving model above. In the critical phase, when $N \sim L$ and $b > 3$ they are asymptotically the same. However, for $b < 3$ they differ as they also do when the number of links does not scale as the number of nodes. The cut-offs do not agree in form, but for $N \sim L$ they agree in quantity.

As detailed in [145], such cut-offs are induced by the disallowal of multiple links and the requirement that the networks be uncorrelated. The absence of multiple links has been put forward by several works as a possible origin for the disassortative correlations found in the Internet and other non-social networks—see for example [56, 146].

5.3.7 Numerical results for disallowal of self links

The non-conserving network model without self links is simulated using a similar algorithm to that presented in Section 5.3.4, but with all events that create a self link suppressed, including the creation of the initial random network configuration.

As with the case with self links allowed, all simulations were run for $\mathcal{O}(10^7)$ Monte Carlo sweeps from random initial conditions of a number of links close to that expected for the stationary state. In order to show all phases with b and k fixed and varying s, $b = 2.25$ and $k = 1.5$ were chosen, along with the following values for s: $s = 2$, to show the low density phase; $s = 1.425$, to show NSL critical sub-phase (a); $s = 1.3125$, to show NSL critical sub-phase (b) (i); $s = 1.125$, to show NSL critical sub-phase (b) (ii); $s = 0.8$, to show NSL critical sub-phase (b) (iii); and $s = 0.3$, to show the high density phase.

The results from the simulations are presented in Figure 5.6 along with theory lines, taken from (5.90), (5.95), (5.102), (5.107), (5.115) and (5.122), for comparison. From the presented results it is apparent that the theory lines do not match up well

for NSL critical sub-phase (b) (i) and NSL critical sub-phase (b) (iii). However, in Figure 5.6 data are presented that compare the network model without self connections with the network model with self connections and the ZRP model; these data seem to indicate that the analysis has at least correctly predicted which phases differ in behaviour between the network models, namely critical sub-phase (b) (i) and critical sub-phase (b) (ii).

A possible explanation for the discrepancy between the theory and simulation data is that the parameters chosen give strong finite-size effects. The values of b and k were initially chosen to allow for a low density phase with a reasonable relaxation time and all other phases to be realised by only varying s. This seems to bring with it values of k and s such that the divergence in the balance equation, i.e L^{k-s}, is very weak and diverges sub-linearly for all but the high density phase. Thus it may be that the divergence for systems with $L = 5000$ nodes, as simulated, is too small for the analysis to be fully valid, i. e. the system may be too small for the asymptotics to be a good enough approximation. This does not explain why critical sub-phase (a) and critical sub-phase (b) (ii) match up well, but there could be something particular to the analysis of these regions that causes them to match up well. Also for finite size systems it would certainly not be unreasonable to expect that the phase boundaries become blurred, as the phases are only strictly distinct in the thermodynamic limit. For these parameters some of the phases and their sub-divisions are quite close together; it could be that some phases are being obscured completely by the finite size of the system.

These parameters do manage to give good data for the high and low density phases, critical sub-phase (a) and critical sub-phase (b) (ii). Particular note should be made of the low-density phase, which agrees excellently with the theory, as this phase was not explored in the data presented for the non-conserving ZRP and network with self links models in previous sections.

5.4 Connection with self-organised criticality

The central ideas of this chapter, i. e. a system that has a critical phase rather than a critical point or that spontaneously drives itself to criticality, are not new. Termed self-organised criticality (SOC), the study of such systems has enjoyed high popularity in years gone by and is still an important, although perhaps less active field, today. For a review of SOC, see [147] for example. Many parallels can be drawn between the ideas central to SOC and the models presented in this chapter.

The models presented above centre around an underlying phase transition—the condensation transition of the ZRP. To this, ingredients are added that allow particles

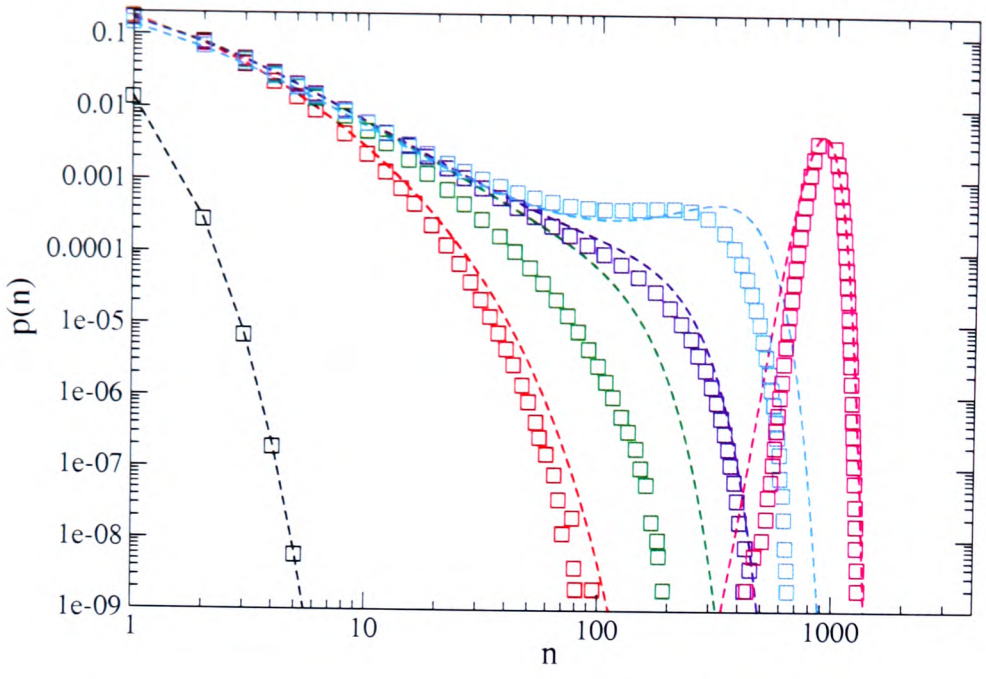


Figure 5.6: Steady state degree distributions from simulations of the non-conserving network model with no self-links allowed (□) compared with theoretical asymptotic curves (dashed lines). Data are shown for: NSL low density phase, $s = 2$ (black); NSL critical sub-phase (a), $s = 1.425$ (red); NSL critical sub-phase (b) (i), $s = 1.3125$ (green); NSL critical sub-phase (b) (ii), $s = 1.125$ (indigo); NSL critical sub-phase (b) (iii), $s = 0.8$ (cyan); NSL high density phase, $s = 0.3$ (magenta).

to enter and leave the system, in such a way that a critical phase is made possible. Connections between SOC and systems with an underlying phase transition have been made before. In [148], a conceptual framework for SOC was given based on the tuning of the order parameter to the critical value in a system with an underlying phase transition. The systems discussed are characterised by a flux which is zero on one side of the critical point and non-zero on the other and thus is an order parameter. By driving this order parameter at an infinitesimal value, it is assured that the system will be at the critical point; this connects the infinitesimal driving often seen in SOC systems with the infinitesimal driving around a critical point of a system with an underlying phase transition. This idea is closely related to the idea behind the models presented above: the ZRP has an underlying phase transition, the condensation transition, and the creation rate at sites can be thought of as the driving rate for the system and is indeed infinitesimal. However, it could also be argued that the overall driving rate, i. e. the sum of all the creation rates, need not be infinitesimal to see the critical phase, although critical sub-phase (a) is not realisable for a non-vanishing overall creation rate; only critical sub-phase (b) is attainable in this case.

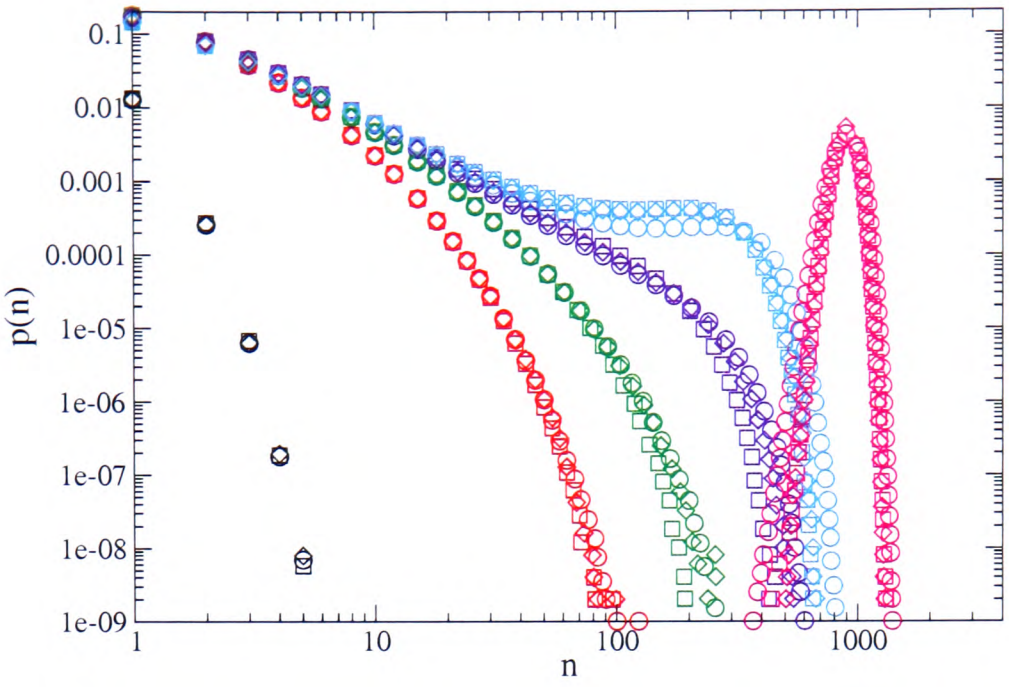


Figure 5.7: Steady state distributions of the degree and the number of particles at a site from simulations of the non-conserving network model with no self-links allowed (□), the non-conserving network model with self-links allowed (◊) and the non-conserving ZRP model (○). Data are shown for: NSL low density phase, $s = 2$ (black); NSL critical sub-phase (a), $s = 1.425$ (red); NSL critical sub-phase (b) (i), $s = 1.3125$ (green); NSL critical sub-phase (b) (ii), $s = 1.125$ (indigo); NSL critical sub-phase (b) (iii), $s = 0.8$ (cyan); NSL high density phase, $s = 0.3$ (magenta).

Another work that discusses the connections between SOC and underlying phase transitions is [149]. In that work, the focus is on systems with an underlying absorbing state phase transition (where the system makes a transition to a state which it cannot subsequently leave) and the local dynamics of the system is then coupled to an external supervisor or a drive. To give SOC, the supervisor or drive indirectly affects the relevant parameter(s) to the phase transition in such a way that the system is attracted to the critical point. Two such ways in which this may occur are infinitesimal driving (where the system is driven by some process at some infinitesimal rate) and extremal dynamics (where the dynamics of the system are such that the constituent with an extreme value of some quantity is the one that changes) The former, in particular, can be used to realise SOC in the following manner: Take a system with an absorbing state phase transition at some global critical density that is conserved by the dynamics; add to this an infinitesimal process that increases the density when the system has reached an absorbing state and an infinitesimal process that decreases the density when in the active state. Due to the infinitesimal nature of the driving the system will move towards the critical point and stay there in the thermodynamic limit. Again this can be

compared with the models presented above, there creation and annihilation dynamics were added to the dynamics of a conserving system. However, the ZRP does not possess an absorbing state phase transition for the rules given above. The argument that the driving rate need not be infinitesimal also applies in this case; the overall annihilation rate need not be infinitesimal either as the creation and annihilation must balance in the stationary state. Also the annihilation rates as given by (5.2) and (5.64) can strictly become finite for $n \geq L$, although this is actually never realised by the dynamics as the largest possible cut-off becomes appreciable for $n \sim L^{k/(k+1)}$. The given annihilation rate can be compared with extremal dynamics: it does favour annihilation from the most occupied site, although it also allows for annihilation from lesser occupied sites with decreasing probability.

Thus, the ZRP and network models presented in this chapter have strong connections with ideas from the field of SOC. Similar models that possess underlying phase transitions and have mechanisms added that drive them to the steady state have already been proposed as examples of SOC and the general idea has been presented as a possible general framework for studying SOC. The models proposed in this chapter are not exactly the same as those proposed in previous work: they have several differences for which it is interesting that the general idea still holds. A further model that links networks and SOC has been proposed in [150]. There a fixed number of nodes are assigned a fitness and links are slowly added to the system at random. When a node receives a link it can become unstable with a probability dependent on its fitness or its degree. An unstable node will lose all its links, either through deletion or rewiring. The rewiring can trigger further instabilities and can cause avalanches of rewiring in the system. Avalanches are a behaviour that is often associated with SOC. Under certain conditions these avalanches can cause the system to have a power-law degree distribution and without fine-tuning of parameters. The dynamics of this model are appreciably different from those presented in this chapter and yet the results are quite similar. It would thus also be interesting to see if the ideas of SOC can be further applied to networks, or a more general model can be formulated.

5.5 Summary

In this chapter, a non-conserving generalisation of the ZRP has been introduced that yields a critical phase instead of the usual critical point in the standard conserving ZRP. This was then related to a network model with annihilation and creation of links and this was shown to possess the same phase diagram; in particular, it also exhibited a critical phase. Thus, a model of a non-growing, rewiring network with

creation and annihilation of links that can have scale-free behaviour in a region of the parameter space has been proposed. This goes some way to fixing the problem of having scale-free behaviour only at a critical point: a problem that is shared by several statistical ensemble models of networks, some of which are discussed in Chapter 4 Section 4.6. However, this model is defined through its dynamics, and these dynamics may not have a realistic interpretation in terms of the restructuring processes that real networks undergo. Nevertheless, the models presented do add positively to the collection of non-growing network models. The success in reproducing a scale-free phase also poses relevant questions as to whether the general idea of these simple models can be applied in models with rewiring dynamics that are more realistic. Some of the results of this chapter have been published in a more concise format in [151] and an in-depth discussion is in preparation.

Chapter 6

Multi-Species Zero-Range Process Model of a Directed, Weighted Network

6.1 Introduction

At the birth of any of field of research, especially in physics, it is common to start with the simplest possible cases, before moving on to study more complicated—perhaps more realistic—cases. So it was in the field of networks: the majority of the models discussed in Chapter 4 were concerned only with the structure of the connections and not really with the specifics of the system they represented. Furthermore, the connections were either present or absent; any possibility of connections being of different type or importance was not considered. It is clear in some networks that the links will differ in importance: in social networks, for example, a link between two best friends has a different significance to a link between two casual acquaintances. In order to treat such differences weighted networks were introduced. In these, each link is assigned a weight to represent its relative importance, use or any other relevant quantity. Following the initial surge of research on unweighted networks in the physics literature, weighted networks have recently become popular [152, 153, 154, 155].

In this chapter, the application of the ZRP to the field of complex networks is continued. Specifically, a multi-species ZRP is used as a model for directed, weighted networks. The steady state of the model can be solved exactly under quite general conditions; this makes the model particularly amenable to analysis. The mapping of the multi-species ZRP to weighted networks is very different than the previous mapping between the single-species ZRP and unweighted networks: sites are still

interpreted as nodes, but particles now carry link and node information. This mapping is not as closely related to previously known models as the previous one discussed in Chapter 4 Section 4.6.1 and thus represents a novel application of a ZRP to the field of networks. The application of the multi-species ZRP to weighted networks also brings to light a new condensation mechanism and this is instrumental in allowing the model to reproduce some of the behaviour seen in real networks. In fact, there are two condensation mechanisms in the multi-species model and each can allow the system to reproduce some realistic behaviour under certain conditions. Before discussing the mapping, weighted networks are discussed and motivated further in the following section.

6.2 Weighted networks

Weighted networks are networks in which each link carries a weight. This weight can represent the relative importance of each link, the amount of traffic that flows over a link or any other relevant quantity. The weight of a link can be important in calculating some properties of a network: for example, in a general transport network the weight of the links can represent the capacity of the link and this is important in calculating maximal flow properties of the network [156]. It might be expected that the links with the greatest weight have the greatest effect in any system; however, it has long been known that for some properties of social networks, weak links are extremely important [157] and cannot be ignored as a good approximation of the system.

Examples of real networks where the weight of the link is important include: the scientific collaboration network (SCN) [40, 41] in which nodes are scientists and they are linked if they have coauthored a paper, with the weight of the link depending on the number of papers that the two have coauthored; the Sardinian inter-municipality commuting network (SMCN) [158], where nodes are municipalities, links are the transport connections between them and the weights represent the number of travellers using these connections; the world-wide airport network (WAN) [43] where nodes represent airports and links represent the existence of direct flights that connect these airports, with the weight of the link representing the total number of seats available between each airport per month; and the metabolic interaction network of E. coli (MINE) [159] where nodes are metabolites and links connect metabolites that are produced from one another in reactions, with the weight of the link representing the flux between the metabolites in these production or consumption reactions. Along with these explicit examples there are many areas in which the generalisation to weighted networks is

useful, in fact nearly all of the areas to which networks can be applied can benefit from the extra information included in the weighted generalisation.

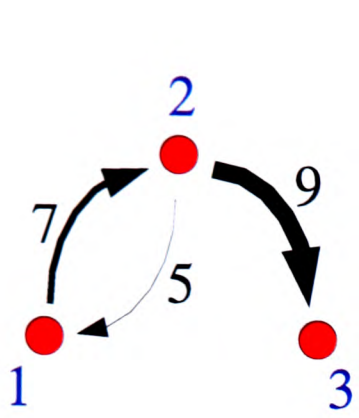
The weight of a link can be represented in many ways, often depending on the specifics of the model. It can be a real number; it can be positive or negative; it can simply be an element of a suitably defined set of possible weights. A simplification that is often used is to represent the weights as positive integers. This allows an interpretation of networks with multiple links as weighted networks, with the number of links between two nodes being taken as the weight between the two nodes. This also allows the weight of a link to be represented by the number of particles associated with a link, which gives the first indication of a possible ZRP mapping.

The structure of a network can be represented by what is known as the adjacency matrix. The rows and the columns of the matrix represent nodes and the entries represent links. For a non-weighted network the entries are 0 and 1 to represent the absence and presence of a link respectively. Thus the (i, j) element of the matrix represents whether a link is present or absent between nodes labelled i and j. For an undirected network the adjacency matrix is symmetric, as there is no difference between a link from node i to j and a link from node j to i. The model that will be introduced is for a directed, weighted network. Thus, the adjacency matrix is not symmetric and the elements can take values other than 0 or 1 to represent the weights of the links. For example, if the i, j element of the adjacency matrix is equal to 5, then this represents a link of weight 5 pointing from node i to node j. The i, j element of the weighted adjacency matrix is denoted by w_{ij}. A simple directed, weighted network is shown in Figure 6.1 along with its adjacency matrix. The adjacency matrix representation proves useful conceptually, both for the model that is introduced in this chapter and for the characteristics that describe weighted networks. For example, the in-degree of node j can be obtained from the weighted adjacency matrix as follows

$$k_j^{\text{in}} = \sum_i \theta(w_{ij}) \,, \tag{6.1}$$

where k_j^{in} denotes the in-degree of node j and $\theta(w)$ is the Heaviside step function [10]. The out-degree distribution is found by reversing the indices on the summand.

As with unweighted networks, simple properties that can be used to characterise the structure of a network are required for weighted networks. Some of these properties carry over directly from unweighted networks. For example, the degree is still the number of connections a node has and the degree distribution is still a useful characteristic of a weighted network. Some properties either lose their meaning in a weighted network and must be generalised, or retain their meaning but with a meaningful generalisation also possible. Some of the key new characteristics for weighted



		to 1	2	3
from	1	0	7	0
	2	5	0	9
	3	0	0	0

Figure 6.1: A simple example of a directed, weighted network (left) and its adjacency matrix representation (right).

networks are as follows:

- **Link weight**—This is simply the value assigned to each link to represent its relevant importance. In terms of the adjacency matrix, the weight of the link between nodes i and j is represented by the value of the (i, j) element. The distribution of weight amongst the links of the network, $\mathrm{P}(w)$, is an important characteristic of the network. One of the first works to discuss the importance of link weights was [157]: in the context of social networks it was found that for some processes the weak links can be important. It should be noted that the weight of a link is referred to as the strength of a link in some of the literature.

- **Node strength**—This is the weighted generalisation of the degree; instead of measuring the number of links connected to a node it measures the total weight that is connected to a node. The distribution of strength in a network, $\mathrm{P}(s)$, is another important characteristic of a weighted network. For a directed network, the strength generalises to in- and out-strengths measuring the sum of the weight of all links pointing to a node and the sum of the weight of all the links pointing from a node, respectively. In terms of the adjacency matrix, the in-strength, s_j^{in}, of node j is given by

$$s_j^{\mathrm{in}} = \sum_i w_{ij} \ , \tag{6.2}$$

with w_{ij} being the (i, j) element of the adjacency matrix. The out-strength of node i is found by reversing the indices of the summand in the above equation. For an undirected network the in- and out-strengths are identical. The concept of strength was first discussed as an important property in the physics literature in [152]. The strength of the node is also referred to as the node strength (when the weight is referred to as link strength) and the total weight of a node. In

this work the link-weight and node-strength terminology of [43] is used as this minimises confusion.

- **Weighted path length**—The weighted path length can be defined in several possible ways. One such way is to assign a relationship between the weight of a link and the cost of traversing it. Thus, instead of being the minimum number of links that must be traversed between two nodes, the weighted path length is the minimum cost that must be traversed between two nodes. This cost could represent, for example, the inverse of the speed or bandwidth of an link in the Internet; thus the path with the least cost may prove to be the fastest possible route to send data by. Weighted path lengths have been discussed in [41, 160].

- **Weighted clustering**—The usual clustering coefficient can still be applied to weighted networks to measure the tendency for highly connected groups to form, but if links have relative importance then there needs to be a measure of the total weight of a cluster, to prevent occurrences such as many weak clusters giving a false view of the clustering of the system. A generalised clustering coefficient has been proposed in [43] with the following form

$$c_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{(w_{ij} + w_{ih})}{2} \theta(w_{ij} w_{ih} w_{jh}) \,, \tag{6.3}$$

thus clusters with high total weight give a greater contribution to the clustering. The weighted clustering is given by the average of the weighted clustering coefficient over all nodes in the network. Note that it is not completely obvious how to best generalise the clustering and weighted clustering to directed graphs.

- **Disparity**—The disparity is a measure of how evenly the strength of a node is distributed amongst its links; the two extremes of this being all links having equal weight and one of the links having a much greater weight than all the others. A quantity that measures this disparity has been proposed in other contexts [161, 162] and is given by

$$Y_2(i) = \sum_j \left[\frac{w_{ij}}{s_i} \right]^2 \,. \tag{6.4}$$

Whether or not certain links dominate the others in importance is clearly an important measure allowed by including the link weight in the description of the network structure. In the case of directed networks, the disparity can be generalised to in- and out-disparities in the obvious way. The disparity has been discussed in the context of weighted networks in [155].

- **Weight-topology correlations**—The relation between the connectivity of the network and the weight can be determined by measuring the correlation between the strength of a node and its degree, i. e. measuring $s(k)$. This provides an insight into the connection between the weighted and unweighted properties of the network. Weight-topology correlations have been discussed in [43].

There are also many other weighted characteristics such as node-node strength correlations and weighted generalisations of the various betweenness properties. The key properties that are studied in the model that is introduced in this chapter are the weight and strength distributions and the usual degree distributions. The disparity and weight-topology correlations are also discussed.

The study of data from weighted network representations of real systems is not yet as abundant as that from unweighted networks. However, some measurements have been taken for the above highlighted characteristics, hinting at the existence of narrow categories of behaviour that many networks fall into as in the case of unweighted networks. The following kinds of behaviour have been observed:

- **Scale-free weight distributions**—In many real systems the distribution of weight amongst the links has been shown to have a heavy-tail, i. e. to decay more slowly than an exponential [158, 159]. From the data available it is often not clear that this is a definite indication of general scale-free behaviour, but gives strong support to this possibility.

- **Scale-free strength distributions**—The distribution of strength amongst the nodes has also shown a heavy-tail in several data sets from real systems [158, 43, 155]. As before, while there is not yet enough data to conclusively show that power-law strength distributions are generically seen, the data that is available gives strong support to the possibility of this.

- **High weighted clustering**—In the SCN the weighted clustering coefficient has been measured [43] and is found to be very close to the unweighted clustering coefficient indicating that there are many stable clusters with the total weight in each cluster being evenly distributed, on average. The weighted clustering has also been measured for the WAN [43] and also the SMCN [158]. For both these networks, the weighted clustering coefficient is found to be higher than its unweighted counterpart. This indicates the existence of a 'rich-club' phenomenon: many of the clusters are formed with highly weighted links.

- **High disparity**—The disparity has been measured in the SMCN [158] and the MINE [159]. In both cases it was found that many nodes had a strength that

was dominated by a small fraction of the connected links. While not measured as being close to its extreme upper value, the values were far enough from the lower extreme to indicate high levels of heterogeneity in how the link weights are distributed for a node with a given strength.

- **Weight-topology correlations**—In the SCN it has been observed that the strength of a node grows linearly with its degree [43], with the constant of proportionality being the average weight indicating an absence of weight-topology correlations. However, in the WAN [43] and SMCN [158], the strength of a node was found to grow super-linearly with its degree. This indicates the presence of weight-topology correlations: links with large weight tend to be connected to nodes with large degree.

In addition to these highlighted weighted properties, some of the unweighted behaviours carry over to weighted networks. Perhaps the most important of these being scale-free degree distributions.

6.3 Weighted network models

In this section, some of the existing weighted network models in the literature are discussed. Many of these models share common ideas with their unweighted counterparts and some are direct weighted analogues of unweighted models. Not all of the major unweighted models have such analogues, although they may appear in time. Prominence is given both to those models that contain ideas relevant to the multi-species ZRP model presented in this chapter and to those that are directly related to the unweighted models detailed in Chapter 4.

6.3.1 Weighted variants of the BA model with fixed weight assignments

One of the first weighted network models was proposed in [152] and is based heavily on the BA model discussed in Chapter 4 Section 4.4.3. Known as the weighted scale-free model (WSF) this is a growing model of an undirected network where at each time step a new node with m links is added and connects to an existing node i with probability

$$\Pi_i = \frac{k_i}{\sum_j k_j} \,. \tag{6.5}$$

This is degree-driven linear preferential attachment, just as in the BA model. The new element is that each link of the new node is assigned a weight that is proportional

to the degree of node to which it is attached. Each node has a fixed total strength at the time of its introduction, so each link is assigned a fraction of this proportional to the degree of the nodes they attach to. The degree and strength distributions are both found be power laws, but surprisingly numerics indicate that they have different exponents. However, an analytical approach indicates that this discrepancy may be due to strong finite size effects and the two distributions should have the same exponent. The weight distribution has a tail that decays more rapidly than an exponential; this is less surprising as the link weight has a maximum possible value and the link weights are fixed after their initial assignment. Various other generalisations of this model are discussed briefly in [152], the most important of these being when links attach to existing nodes with a probability dependent on their strength, rather than on their degree. This concept is also important in one of the weighted network models that follows, as well as the multi-species ZRP network model introduced in this chapter.

A related model where links connect to a node with a probability proportional to its strength, but for which the weight assignments are random has been studied in detail in [163]. In that work it was found that the node strength has a power-law distribution asymptotically with an exponent equal to 3; this result is independent of the link weight distribution.

6.3.2 Weighted variants of the BA model with evolving weights

The Barrat-Barthélemy-Vespignani (BBV) model, first introduced in [164], is another model that is strongly related to the BA model and its basic attachment mechanism is that of the WSF model with strength driven attachment. However, the BBV model adds a new element in that the weights of the links are allowed to change in time. Specifically, whenever a node receives a new link, the weight of its existing links are also updated.

The details of the model are as follows. Starting from an initial seed of m_0 nodes, at each time step a new node with $m < m_0$ links is added to the network. Links from this node are connected to existing nodes of the network with a probability that depends on the strength of the node. Thus a link is connected to node i with probability

$$\Pi_i = \frac{s_i}{\sum_j s_j} , \tag{6.6}$$

i. e. new connections are governed by strength driven linear preferential attachment. Each new link has a weight w_0 at its inception. When a new link is attached to a node, i, the weights of the links already connected to node i (w_{ij}) are increased by an

amount

$$\delta \frac{w_{ij}}{s_i}\,, \tag{6.7}$$

where δ is the total weight increase due to each new link. Thus the added weight is distributed amongst the existing links of node i in proportion to the weight of each existing link. The strength of node i increases by an amount $\delta + w_0$; this accounts for the increase of the weight of the existing links and the weight of the newly added link.

A detailed analysis of this model and several of its generalisations is presented in [165] and a directed version is discussed in [166]; the relation of results from the model with real data is discussed in [155]. Some interesting points are that in its basic form the model displays: power-law degree, weight and strength distributions with the same exponent for the degree and strength distributions; high clustering, with the weighted clustering coefficient appreciably larger than its unweighted counterpart; and while the strength grows linearly with degree, the constant of proportionality is not $\langle w \rangle$ indicating the presence of weight-topology correlations.

In a similar model, proposed in [154], new links attach to nodes by attaching to the end of a link which is chosen with a probability proportional to its weight; this link also has its weight increased. This model also gives rise to power-law degree, weight and strength distributions; the degree and strength distributions again share the same exponent.

Both the above models have the basic underlying ideas that strong nodes should attract more links and hence more strength, and that weighty links should attract more weight in the evolution of the network. These two concepts—that strong nodes attract more strength and weighty links attract more weight—are central to the multi-species ZRP model that is introduced later in this chapter.

Another related model has been proposed in [167]. In this model, the new nodes connect with the degree-driven linear preferential attachment of the BA model. In addition, at each time step a number of links are chosen with a probability proportional to their weight and each has its weight increased by a fixed amount. This model gives the same scale-free degree behaviour as the BA model but also scale-free weight and strength distributions. The model can also show a lack or presence of weight topology correlations depending on the ratio of newly added links to links that increase their weight, at each time step.

A further model that grows according to the BA model with an additional attractiveness, i. e. links attach with a probability $\Pi_i \propto k_i + \lambda$, has been proposed in [168]. Here weights are assigned according to the degree at either end of the link. The link between nodes i and j has a weight $(k_i k_j)^\theta$. Thus as nodes gain new links the weights

of existing links evolve. The model is found capable of giving power-law degree, weight and strength distributions.

6.3.3 Static weighted network models

Thus far, generalisations of the BA model, which is a network model that evolves through growth, have been discussed. There also exist models that generalise some of the static, non-growing models of networks and these are discussed in this section.

A weighted generalisation of the configuration model has been proposed in [154]. Recall that the original conflguration model was, in essence, a maximally random network under the constraint of a given degree distribution. In the generalisation presented in [154] the same idea is used with the addition of a constrained weight distribution. The resultant strength distribution under these constraints was calculated. It was found that if either of the degree or weight distributions is a power law with the other rapidly decreasing, or if both the degree and weight distributions have power-law forms, then the resultant strength distribution has a power-law form. The exponent of the strength distribution in these cases was found to be equal to the exponent of the power-law of the underlying degree or weight distribution; in the case where both have power-law forms, the strength exponent was the smallest of the two.

A separate generalisation of the configuration model to weighted networks was proposed in [169]. In the original configuration model each node is given a number of stubs, i. e. ends of links, drawn from the degree distribution; the stubs are then randomly paired to produce links. This allows for multiple and self-links, although the number of these occurrences is expected to be low, for scale-free degree distributions with bounded fluctuations. In the generalisation proposed in [169], scale-free degree distributions with unbounded fluctuations are studied. In this case a large number of multiple links are generated and each of set of these is then taken to be a single link with a weight equal to the number of multiple links in the set. Thus, the initial distribution has become that of the strength rather than the degree. It was found that the degree behaved as a power-law with exponent equal to 2 even though the exponent of the strength distribution is less than 2. Thus for this model there is a non-trivial scaling relation between the strength and the degree.

6.3.4 Weighted network fitness models

A weighted generalisation of one of the multiplicative fitness variants of the BA model has been proposed in [167]. In this model, at each time step a new node is added with

m links and the links of this node attach to existing nodes with probability

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j} , \tag{6.8}$$

where η_i is the fitness of node i and is chosen from some distribution, $\rho(\eta)$, at the inception of the node. This is the multiplicative fitness attachment rule from before, see Chapter 4 Section 4.4.4. For this weighted version each link is taken to have a fixed weight, w_0, when it is created. A fitness, ζ_{ij}, is also assigned to each link when it is created, from a distribution, $p(\zeta_{ij}|\eta_i, \eta_j)$, that can depend on the fitnesses of the nodes at each end of the link. Concurrently with the growth process, at each time step the weight of m' chosen links increases by w_0; link i, j is chosen with probability

$$\Pi_{ij} = \frac{\zeta_{ij} w_{ij}}{\sum_{l,l'} \zeta_{l,l'} w_{l,l'}} . \tag{6.9}$$

Thus, as the network grows, well connected and fit nodes gain more links and weighty and fit links gain more weight.

This model displays a rich phase behaviour. It is capable of generating networks with or without weight-topology correlations. It can also display Bose-Einstein condensation of links onto nodes and weight onto links where the fittest nodes and links capture a finite fraction of the available links and weight respectively. The model is also capable of displaying the extremes of disparity behaviour, as may be expected in a model that can display link and weight condensation.

6.3.5 Other weighted models

There are many other weighted networks in existence and the number is growing all the time. One such model is introduced in [170], where new nodes are added at each time step and attach to nodes with a probability dependent on the amount of traffic that would flow through it in an idealised system. In this system nodes all transmit equal amounts of data at regular intervals and data is always sent via the shortest possible route. The weight of each link is taken to be the amount of traffic that passes across it this idealised transport system. Thus the weights of the links update according to non-local rules: the addition of new nodes can cause the shortest paths to change and these determine the amount of traffic that passes over each link. This model can produce power-law strength and degree distributions with strong weight-topology correlations. The general weight distribution due to idealised transport systems in weighted networks has been studied in [171].

Another such model that mixes the ideas of the WSF and fitness models discussed above has been presented in [172]. Here the model grows according to degree-driven

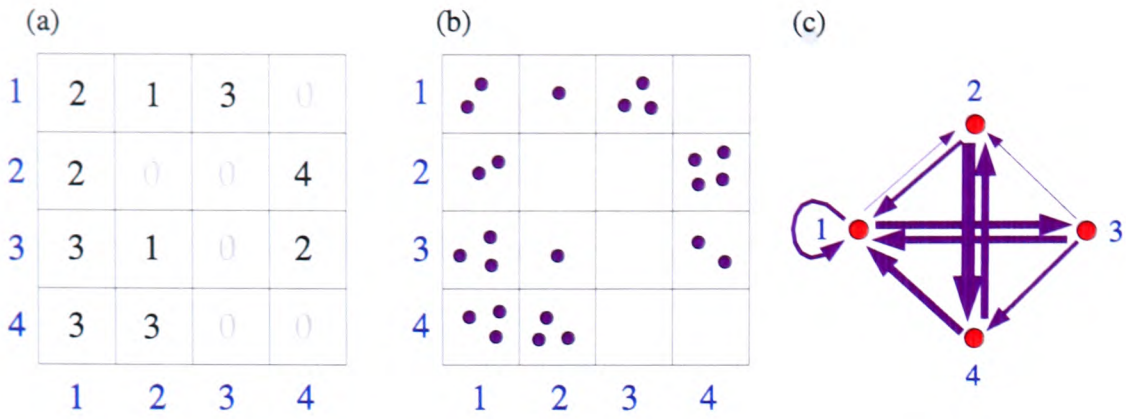


Figure 6.2: A simple example of the mapping between (a) the adjacency matrix of a directed, weighted network and (b) particles on a lattice. Also shown is (c) the corresponding directed, weighted network with the thickness of the lines representing the weight of the links.

linear preferential as for the BA and WSF models. With probability p a weight is assigned to a new link as in the WSF model and with probability $1-p$ the new link has its weight assigned according to the fitness of the node it has attached to. This model can also be generalised to assign weights according to the fitness model discussed in the previous section. This model was seen to show power-law degree and strength distributions and exponentially decreasing weight distributions.

6.4 Weighted network ZRP mapping

The weighted nature of the networks considered in this chapter actually provides new possibilities for mappings to particle-lattice models. Earlier in this chapter the adjacency matrix representation of a weighted network was introduced. In that representation the weight of each link is represented by the relevant element of the adjacency matrix, with a weight of zero indicating the absence of a link, as depicted in Figure 6.1. If the weight of a link is restricted to being a positive integer, then the adjacency matrix itself can be represented by a two-dimensional lattice with a number of particles on it: the number of particles at a site represents the entry in the corresponding adjacency matrix. A simple example of this is shown in Figure 6.2.

In fact, any model that deals with the placement of particles on a two-dimensional lattice with no exclusion rules can be thought of as a model of a weighted network on some level: the configuration of the particles of such a model can always represent the configuration of a weighted network. A similar claim was made for unweighted networks in Chapter 4 Section 4.7. However, it should be noted that this mapping is quite different from that case; in many ways it is actually superior. This mapping is now a

direct one, each configuration uniquely defines a directed weighted network—only if the configuration of particles is guaranteed to be symmetric about the lattice diagonal will an undirected network be described. This contrasts with the previous mapping, as there each particle configuration in principle corresponded to multiple possible network configurations. Thus the mapping between two-dimensional lattice models and directed weighted networks avoids many of the subtleties involved with correctly calculating the network ensemble properties from configurations of the particle-lattice model, as discussed in Chapter 4 Section 4.7. Another difference is that if the number of particles is conserved then only the total weight of the corresponding network is conserved under the dynamics; extra constraints would have to be introduced to fix, for example, the number of links in the network.

This mapping between two-dimensional particles-on-lattices models and directed, weighted networks can be used to give either: static models, where the particles are placed once and then fixed; or rewiring models that generate ensembles, where the particles are continually rearranging themselves. It would perhaps be difficult, but in principle there is no reason why a growing lattice, and hence a mapping to a growing weighted network, would be impossible. As before a ZRP network model produces a rewiring network that generates an ensemble of networks as the particles are continually moving and hence the network is continually changing.

One way of representing a directed, weighted network in this way is to use a single-species ZRP. The rules of the ZRP mean that in the corresponding network single units of weight would rewire themselves in a way that depends on the weight of the link that they rewire away from. A simple example of such a move is shown in Figure 6.3. If the hop-rate for the ZRP is chosen to decrease with increasing particle number, then links with high weight will lose weight more slowly; this is equivalent to weighty links attracting more weight—one of the principles behind some of the weighted network models discussed above in Section 6.3. In particular the choice $u(n) = \beta(1 + b/n)$ will produce a network that can have a power-law weight distribution at a critical global density of weight. Above the critical density, the network will be in a condensed phase where a single link will take up a finite fraction of the available weight. Below the critical density the weight will be distributed among the links according to an exponential distribution.

There also exists a mapping of the multi-species ZRP to weighted networks; one that can treat the ideas that weighty links attract more weight and that strong nodes attract more strength. This mapping is discussed in detail in the following section.

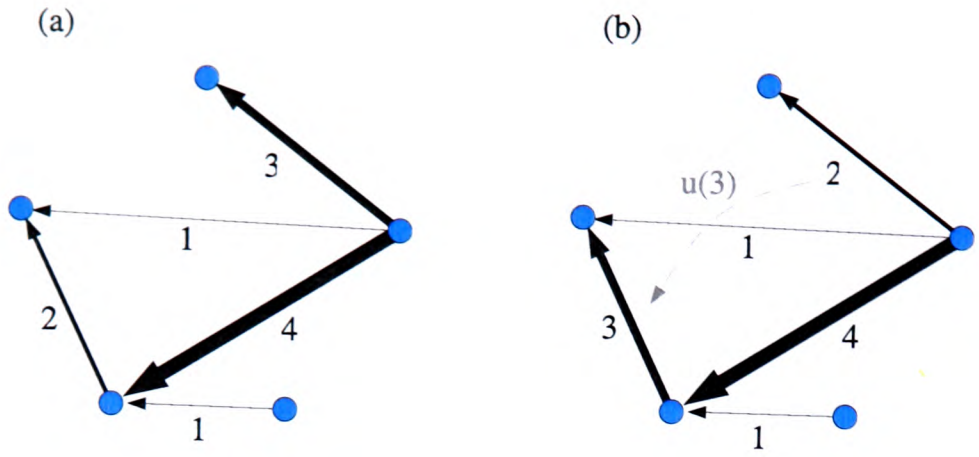


Figure 6.3: An example of a possible rewiring of a single unit of weight under the rules of a basic ZRP model. The network is shown before (a) and after (b) the move. The move itself is shown in grey and happens with a rate $u(3)$ as this was the weight that the link being rewired away from had before the move. In this model the unit of weight can rewire to a link that shares no ends with the link being rewired from.

6.5 Multi-species ZRP mapping to directed, weighted networks

It has previously been discussed how a ZRP can be defined with multiple species of particle which can interact with each other and still give a system for which the steady state can be solved—see Chapter 2 Section 2.5.2. This allows for a more complex model of weighted networks that can treat both weight and strength interactions.

The key point of the mapping is that the weighted adjacency matrix and corresponding two-dimensional lattice can be represented by a one-dimensional lattice with multiple species of particle. Each column (row) of the lattice can be represented by a distinct site and each row (column) of the lattice can be represented by a distinct particle species. Thus there are as many species of particle as there are nodes in the network and sites of the lattice. An example showing this equivalence is given in Figure 6.4, along with the corresponding directed weighted network. In the network context a particle of the multi-species ZRP represents a unit of weight; the site at which it is located represents the node to which it is pointing; and its species represents the node from which it is pointing. The meaning of the site and species can be reversed, but the above meaning is taken in this chapter for simplicity.

Within this mapping the total number of particles at a site represents the in-strength of the corresponding node. The total number of particles of a species represents the out-strength of the corresponding node. Now, in the multi-species ZRP, particles move on the one-dimensional lattice with a rate that can depend on the

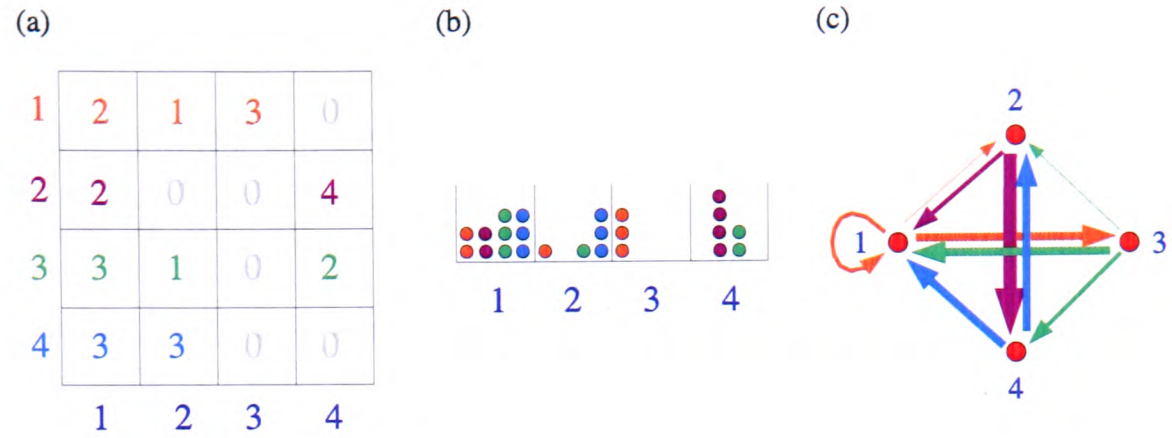


Figure 6.4: An example of the equivalence between adjacency matrix of a directed, weighted network (a) and a multi-species ZRP (b). Each particle represents a unit of weight of a link. The species of a particle represents where the link is pointing from and is distinguished by the colour of the particle. The site at which the particle resides represents where the link is pointing to. Also shown is the corresponding directed weighted network (c), with the thickness of the lines representing the weights of the links. The adjacency matrix and the network have also been colour coded to clarify the equivalence.

number of particles of its own species and on the numbers of particles of all the other species. This general case includes having an explicit dependence on the number of particles of the same species at a site and the total number of particles at a site. Hence, in the network context units of weight rewire with rates that depend on the in-strength of the node being rewired from and the weight of the particular link that is being rewired; such a move is depicted in Figure 6.5. When a unit of weight is rewired, the node to which it is pointing changes but the node from which is was pointing stays the same, i. e. units of weight are rewired from their target end. Under these rules the number of particles of a species is fixed, hence the out-strength of each node is fixed. The total number of particles at a site or in-strength, however, is allowed to vary. Thus the multi-species ZRP represents a network model where the out-strength of each node is fixed and the in-strength varies by single units of weight changing which node they point to under rules which can depend on the weight of the link being rewired and the in-strength of the node being rewired from. It should be noted that for this mapping the directed nature of the network is important. For an undirected network the adjacency matrix would have to be symmetric and this would introduce extra constraints on the allowed hopping of the particles which would remove the zero-range nature of the model.

Thus the key features of the mapping of the multi-species ZRP to a directed,

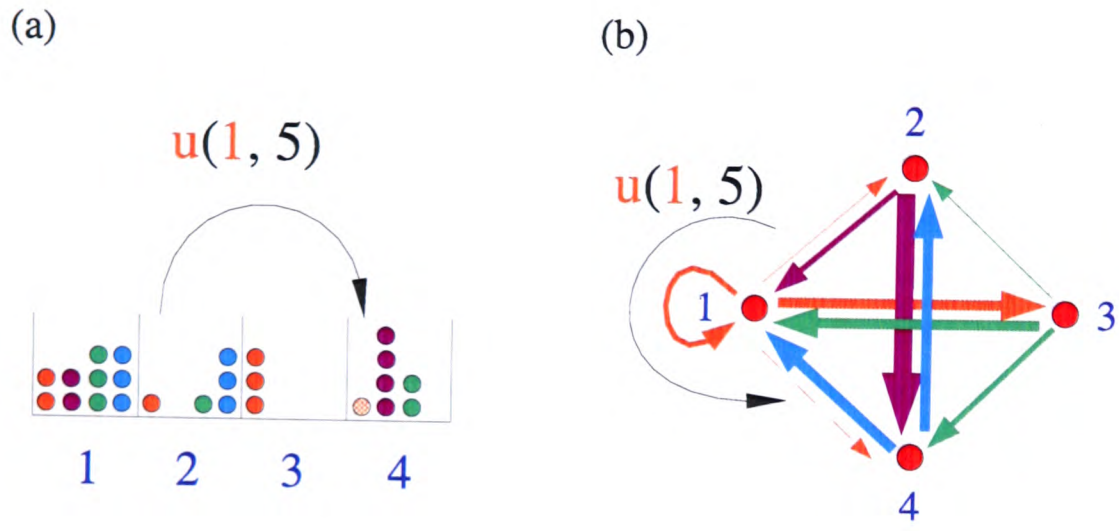


Figure 6.5: An example of a rewiring move in the multi-species ZRP model of a directed weighted network. (a) The move in the context of the multi-species ZRP; an orange particle at site 2 hops to site 4 with a rate that depends on the number of orange particles at site 2 (=1) and the total number of particles at site 2 (=5). (b) The move in the context of the directed, weighted network (note that the thickness of the lines represents the weight of the link); a unit of weight—in this case a whole link as it only has weight 1—pointing from node 1 to node 2 rewires to point to node 4 with a rate that depends on the weight of the link from which a unit of weight is rewiring (=1) and the in-strength of the node which the link is pointing to prior to rewiring (=5).

weighted network are

- Particles represent units of weight.
- The species of a particle represents which node the relevant link is pointing from.
- The site at which a particle resides represents which node the relevant link is pointing to.
- The number of particles at a site corresponds to the in-strength of the relevant node.
- The number of particles of a given species corresponds to the out-strength of the relevant node.
- The ZRP model evolves by single particle hops from site to site which occur with a rate that can depend on the number of particles of the same species and the total number of particles at the site that is being hopped from.
- The corresponding network model evolves by single units of weight rewiring to point towards a different node than before with a rate that can depend on the weight of the link being rewired and the in-strength of the node being rewired from. Thus the out-strength of each node is conserved under the dynamics, but the in-strength can evolve.

6.6 Simple model of weighted networks

The steady state for the multi-species ZRP can be solved under quite general conditions, more general even than those discussed above, and can be further generalised in certain ways and retain a soluble steady state. Some of these conditions and generalisations are discussed in Section 6.8. For simplicity, the specialised version of the multi-species ZRP briefly introduced above will be discussed in detail; this version is perhaps the simplest that models directed, weighted networks and produces non-trivial behaviour. Where possible, the model will discussed in both terms of the multi-species ZRP and the corresponding directed, weighted network interpretation.

6.6.1 Model definition

The model is defined on a lattice of L sites upon which reside L different species of particle. The number of particles of species k on site l will be denoted n_{kl}; this represents the weight pointing from node k to node l and is equivalent to the k, l

element of the weighted adjacency matrix. The number of particles of species k is denoted N_k and is equivalent to the out-strength of node k. The total number of particles at site l is denoted by X_l, is given by $X_l = \sum_k n_{kl}$ and is equivalent to the in-strength of node l. Note that the labellings L for the number of sites of the lattice and N for the number of particles have been retained from the ZRP terminology, where L stands for length and N for number. This is unfortunate in the networks context as L does not stand for the number of *links* and N does not stand for the number of *nodes*, instead the number of nodes is labelled L and the number of links is labelled K. This is also different from the unweighted network to ZRP mapping as there N was the number of link ends or twice the number of links.

The particles move on the lattice under the following dynamical rules. A particle of species k hops from a site l to another randomly chosen site, i. e. the geometry is fully connected, with a rate u, which depends on the number of particles of the same species and the total number of particles at the site. The model discussed here specialises to the following factorised form of the hop rate

$$u\left(n_{kl}, X_l\right) = u_w(n_{kl})\, u_s(X_l)\, \theta(n_{kl}) \;, \tag{6.10}$$

where $\theta(n)$ is the usual Heaviside step function [10], $u_w(n_{kl})$ depends solely on the number of particles of species k at site l and $u_s(X_l)$ depends solely on the total number of particles at site l. In the network context, u is the rewiring rate and this factorises into u_w which depends on the weight of the link that is to be rewired and u_s which depends on the in-strength of the node that is being rewired from, hence the w and s labellings. More general forms for the hop rate are discussed in Section 6.8.

6.6.2 Steady state

With the above factorised form for the hopping rate the steady state is quite straightforward to prove; the usual conditions on the multi-species hop-rates for a factorised steady state (2.42), as discussed in Chapter 2 Section 2.5.2, are automatically satisfied when the hop rate factorises in this way. As a further simplification, the number of particles of each species is assumed to be the same; this lightens the notation as N_k can be written as N. In the network context this corresponds to fixing the out-strength of each node to be the same. Under these conditions the probability of a configuration in the steady state can straightforwardly be proven from first principles and this is done as follows. As an ansatz, the steady state probability of a given particle configuration $\{n_{kl}\}$ is assumed to have the following factorised form

$$\mathrm{P}\left(\{n_{kl}\}\right) = \frac{1}{Z(L,N)} \prod_{l=1}^{L} \left[f_s(X_l) \prod_{k=1}^{L} f_w(n_{kl}) \right] \;, \tag{6.11}$$

where

$$f_s(X) = \begin{cases} \prod_{\mu=1}^{X} \frac{1}{u_s(\mu)} & \text{for } X > 0 \\ 1 & \text{for } X = 0 \end{cases} \tag{6.12}$$

$$f_w(n) = \begin{cases} \prod_{\mu=1}^{n} \frac{1}{u_w(\mu)} & \text{for } n > 0 \\ 1 & \text{for } n = 0\,, \end{cases} \tag{6.13}$$

and $Z(L,N)$ is a normalisation constant given by

$$Z(L,N) = \sum_{\{n_{kl}\}} \prod_{l=1}^{L} \left[f_s(X_l) \prod_{k=1}^{L} f_w(n_{kl}) \right] \prod_{k=1}^{L} \delta\left(\sum_{l=1}^{L} n_{kl} - N \right) . \tag{6.14}$$

This is a sum over all possible states of the un-normalised probability. The L delta functions ensure that only those terms of the sum that have the correct number of particles of each species contribute.

As the model is defined on a fully-connected geometry, it is possible that a solution exists which satisfies the detailed balance condition

$$\begin{aligned} &\mathrm{P}(n_{11},\ldots,n_{kl},\ldots,n_{\mu\nu},\ldots n_{LL}) u_w(n_{kl}) u_s(X_l) = \\ &\qquad \mathrm{P}(n_{11},\ldots,n_{kl}-1,\ldots,n_{\mu\nu}+1,\ldots n_{LL}) u_w(n_{\mu\nu+1}) u_s(X_\nu+1)\,, \end{aligned} \tag{6.15}$$

i. e. the current out of a configuration $(n_{11},\ldots,n_{kl},\ldots,n_{\mu\nu},\ldots n_{LL})$, labelled i say, into a configuration $(n_{11},\ldots,n_{kl}-1,\ldots,n_{\mu\nu}+1,\ldots n_{LL})$, labelled j say, is equal to the current out of configuration j back into configuration i. Thus a solution which satisfies this condition is sought. As stated in Chapter 2, it is known that an irreducible Markov Process with a finite state space possesses a unique steady state, see, for example, [11, 12]. The multi-species ZRP falls within the class of Markov processes as its future evolution depends only on the present state of the system. The irreducible condition—i. e. any state can be reached from any other, given sufficient time—is trivially satisfied when the model is defined on a fully-connected geometry and the hop rates do not go to zero for any non-zero number of particles. Hence, if a solution satisfying detailed balance can be found, then it is the unique steady state of the system. Of course, in the thermodynamic limit the system will have an infinite state space, but it is expected that the argument will generalise to an infinite system in a similar way to the single species system [13] and some discussion of this has been given in [22]. Generalisations with different geometries where the detailed balance condition may not be applicable are discussed in Section 6.8. Cancelling common factors from

6.15 and rearranging a little yields

$$\frac{f_w(n_{kl})}{f_w(n_{kl}-1)}u_w(n_{kl})\frac{f_s(X_l)}{f_s(X_l)-1}u_s(X_l) = \frac{f_w(n_{\mu\nu}+1)}{f_w(n_{\mu\nu})}u_w(n_{\mu\nu}+1)\frac{f_s(X_\nu+1)}{f_s(X_\nu)}u_s(X_\nu)\ . \quad (6.16)$$

As each side of the above equation depends on variables which can be treated as independent, the LHS and RHS can be set equal to a constant and without loss of generality this constant can be set equal to 1,

$$\frac{f_w(n_{kl})}{f_w(n_{kl}-1)}u_w(n_{kl})\frac{f_s(X_l)}{f_s(X_l)-1}u_s(X_l) = 1\ . \quad (6.17)$$

The above will be satisfied when each of the weight and strength dependent factors separately equate to 1. The resultant equations are

$$f_w(n_{kl}) = \frac{f_w(n_{kl}-1)}{u_w(n_{kl})} \quad (6.18)$$

$$f_s(X_l) = \frac{f_s(X_l-1)}{u_s(X_l)} \quad (6.19)$$

which can be solved recursively to give (6.12) and (6.13). Thus with the factorised form of the hop rate (6.10), the system has a factorised steady state given by (6.11).

6.6.3 Condensation theory

An interesting feature of this model is its ability to exhibit two types of condensation transition. One is an independent condensation of each species of particle, with each condensate located at a randomly chosen site. In the network context this is a condensation of the weight onto a link pointing from each node and is thus a realisation of a transition in the disparity. The other is a collective condensation of all species of particle on the same site. In the network context this is a condensation of the in-strength onto a single node, i. e. all nodes point to this one with links that have a finite fraction of the available weight. The exactly soluble, factorised nature of the steady state probability distribution (6.11) allows for the analysis of these condensation transitions and it is this that is discussed in this section.

The steady state as presented in (6.11) is within a canonical formalism, i. e. the number of particles of each species is fixed. As is often the case, it turns out to be easier to work in a grand-canonical formalism where only the the average number of particles of each species in the system is fixed: the actual number of particles of each species at any given time is allowed to fluctuate around this value. There are several paths to a grand-canonical formalism, the simplest is to introduce fugacities

appropriately and then choose them to fix the correct average particle number for each species. In general a separate fugacity is needed for each species. Thus the grand-canonical normalisation is given by

$$\mathcal{Z}(L,\{z_k\}) = \sum_{\{n_{kl}=0\}}^{\infty} \prod_{l=1}^{L} \left[f_s(X_l) \prod_{k=1}^{L} f_w(n_{kl}) \right] \prod_{k=1}^{L} z_k^{\sum_l n_{kl}} \,. \tag{6.20}$$

and as usual this quantity contains much information and is a key component in analysing the condensation transitions. In going from the canonical normalisation (6.14) to the grand-canonical normalisation (6.20) the number of particles of each species at each site is summed from zero to infinity and the delta functions are replaced by fugacities raised to the power of the total number of particles of the relevant species. These fugacities are then chosen to fix the average number of particles of each species to the correct values, i. e. to satisfy the equation

$$z_k \frac{\partial \ln \mathcal{Z}(L,\{z_k\})}{\partial z_k} = \left\langle \sum_l n_{kl} \right\rangle = N_k \,. \tag{6.21}$$

In the above the angle brackets indicate the average taken in the grand-canonical ensemble. The expression for the steady state probability distribution in the grand canonical ensemble is

$$\mathrm{P}\left(\{n_{kl}\}\right) = \frac{1}{\mathcal{Z}(L,\{z_k\})} \prod_{l=1}^{L} \left[f_s(X_l) \prod_{k=1}^{L} f_w(n_{kl}) \right] \prod_{k=1}^{L} z_k^{\sum_l n_{kl}} \,. \tag{6.22}$$

From this, forms for properties such as the average number of particles at a site both for a specific species and all species can be calculated.

As the number of particles of each species has been restricted to be the same, i. e. $N_k = N$, and the system is homogeneous, i. e. all sites are the same statistically, the above expressions (6.20), (6.21) can be simplified. Only a single fugacity is needed; one that is repeated in the expression for each species of particle, i. e. $z_k = z$ for all k. Noting that

$$\prod_{k=1}^{L} z^{\sum_l n_{kl}} = \prod_{l=1}^{L} \prod_{k=1}^{L} z^{n_{kl}} \,, \tag{6.23}$$

the grand-canonical normalisation can be re-written as

$$\mathcal{Z}(L,\{z_k\}) = \prod_{l=1}^{L} \left[\sum_{n_{1l}=0}^{\infty} \sum_{n_{2l}=0}^{\infty} \cdots \sum_{n_{Ll}=0}^{\infty} f_s(X_l) \prod_{k=1}^{L} f_w(n_{kl}) z_k^{n_{kl}} \right] \,. \tag{6.24}$$

Then due to the homogeneity of the system, the normalisation reduces to a site term raised to the power L.

$$\mathcal{Z}(L,z) = \left[\sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \sum_{n_L=0}^{\infty} f_s(X) \prod_{k=1}^{L} f_w(n_k) z^{n_k} \right]^L \,, \tag{6.25}$$

where

$$X = \sum_k n_k \,, \tag{6.26}$$

and the n_{kl} have been relabelled as n_k due to the statistical equivalence of sites. This factorisation of the normalisation constant is analogous to the factorisation of the partition function of a Boltzmann system in the grand-canonical ensemble. However, due to the extra constraints in this multi-species system the factors themselves are more complex and the system allows for non-trivial behaviour. Defining the function $F(z)$ as

$$F(z) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_L=0}^{\infty} f_s(X) \prod_{k=1}^{L} f_w(n_k) z^{n_k} \,, \tag{6.27}$$

the grand-canonical normalisation is written as

$$\mathcal{Z}(L,z) = [F(z)]^L \,. \tag{6.28}$$

The condition on the fugacity to give the correct average number of particles of each species in the system is then written as

$$\langle N \rangle = z \frac{\partial \ln F(z)}{\partial z} \,. \tag{6.29}$$

This should be contrasted with the equation for the basic zero-range process (2.12) discussed in Chapter 2. There a similar logarithmic derivative yielded the average number of particles at a site; here it yields the average number of particles of a species, i. e. L times the average number of particles at a site.

Equation (6.29) can be written in a form that depends on the global density of particles of each species, $\rho = N/L$,

$$\rho = \frac{z}{L} \frac{F'(z)}{F(z)} \,. \tag{6.30}$$

This is now in a form that can more obviously be treated in the thermodynamic limit, i. e. with the number of particles of each species and the number of sites going to infinity, with the global density of particles of each species fixed and finite. Note that (6.30) should be contrasted with (2.14) from Chapter 2 Section 2.3 as it does not have quite the same form in that there is no sum over sites in (6.30). As with the basic ZRP, it is this density equation (6.30) that is the key in the analysis of condensation transitions. If there is some finite critical density, ρ_c, above which equation (6.30) cannot be satisfied, then this signals a condensation transition. As in the case of the basic ZRP condensation analysis, the fugacity z must take a value less than some radius of convergence. If at this maximum value of the fugacity the RHS of (6.30) converges then there will indeed be a finite critical density above which the equation cannot be

satisfied and a condensation transition. The critical density will be equal to the RHS of (6.30) evaluated at the maximum value of the fugacity. A condensation transition is not possible when the RHS of (6.30) diverges as this system is homogeneous.

Note that for the RHS of (6.30) to converge $F'(z)/F(z)$ must be $\mathcal{O}(L)$. This is different from the basic ZRP case, as can be seen by comparing with equation (2.14) where $F'(z)/F(z)$ is summed over sites. In the case of the present directed weighted network model, the form of $F(z)$, (6.27), means that $F'(z)/F(z)$ can take be $\mathcal{O}(L)$ with appropriate choices of the hop rate as is seen in the following analysis.

The two types of condensation are most easily seen in the rather extreme cases of setting either u_s or u_w equal to 1, i. e. causing particles only to interact with only their own species, or with all species equally. In the network context, this corresponds to setting either of the effective strength or weight attractions to zero. The case where both interactions are present is difficult to treat analytically; numerical treatment of this case is presented in Section 6.7.

Independent species condensation

The case where each species condenses independently of the others is most straightforwardly realised when the intra-species interaction is not present, i. e. when $u_s(X)$ is set equal to 1. With this choice, $f_s(X) = 1$ and $F(z)$ factorises, having the form

$$F(z) = \left[\sum_{n=0}^{\infty} f_w(n) z^n\right]^L . \tag{6.31}$$

As the intra-species interaction is not present the system is simply L uncoupled single-species ZRPs; the above form is what is expected for this and can be compared with the basic ZRP case discussed in Chapter 2. With this simple form for $F(z)$, the density equation (6.30) takes on the form

$$\rho = \frac{\sum_{n=0}^{\infty} n f_w(n) z^n}{\sum_{n=0}^{\infty} f_w(n) z^n} , \tag{6.32}$$

which has the same form as equation (2.14) for the basic ZRP, although in this case the equation is explicitly homogeneous. As with the basic ZRP if both the numerator and denominator of the above equation converge at the maximum value of z, then there will be a finite critical density, ρ_c, above which the system will be in a condensed state. Each species will condense independently onto a randomly located site which will hold all the excess particles of a particular species in the system, i. e. it will hold $(\rho - \rho_c)L$ particles of the relevant species. There is nothing to stop more than one species condensing onto the same site, but they do so independently. Due to the

density equation (6.32) having the same form as the density equation for the basic ZRP (2.14), the analysis determining the conditions on the hop-rates required to give condensation is the same as for the basic ZRP as presented in Appendix A. It is the asymptotic form of the hop rate that determines the convergence properties of the RHS of (6.32). The following asymptotic form of the hop-rate will give a system that can undergo condensation

$$u_w(n) \sim \beta_w(1 + b_w/n) \,, \tag{6.33}$$

where β_w is a finite constant and with $b_w > 2$. This gives a maximum value of z equal to β_w. In fact the most direct way to see this is to find the form of $f_w(n)$ that allows convergence and then infer the hop rate from this [4]. It is straightforward to see that an $f_w(n)$ that behaves asymptotically as $f_w(n) \sim \beta_w^{-n} n^{-b_w}$, which is indeed satisfied if the hop rate behaves as $u_w(n) \sim \beta_w(1 + b_w/n)$, guarantees convergence of the RHS of (6.32).

As the form for $F(z)$ has simplified for the case $u_s = 1$, the forms for $\mathcal{Z}(L, z)$ and $\mathrm{P}(\{n_{kl}\})$ are also simple. The normalisation (6.20) simplifies to

$$\mathcal{Z}(L, z) = \prod_{k=1}^{L} \prod_{l=1}^{L} f_w(n_{kl}) z^{n_{kl}} \,, \tag{6.34}$$

and the steady-state probability distribution of a configuration (6.22) simplifies to

$$\mathrm{P}(\{n_{kl}\}) = \frac{1}{\mathcal{Z}(L, z)} \prod_{k=1}^{L} \prod_{l=1}^{L} f_w(n_{kl}) z^{n_{kl}} \,. \tag{6.35}$$

Thus the probability of the number of particles of a single species at a site is given by

$$\mathrm{p}(n) = \frac{f_w(n) z^n}{\sum_{m=0}^{\infty} f_w(m) z^m} \,, \tag{6.36}$$

where this is the same for all sites and species due to the homogeneous nature of the model. The above equations are only fully valid when the system is in the un-condensed state. However, in the condensed state equation (6.36) evaluated at the maximum value of the fugacity will give the distribution of the number of particles of a species at sites where that species is not condensed. This is useful for comparison of the theory with numerical simulation. If the system is in a condensed state the theory line should match the part of the particle occupancy distribution that does not represent the condensate. The condensate appears as a peak at high occupancy that is additional to the background distribution. Equation (6.36) is also useful to determine which hop rates will give a power-law distribution of weight at criticality, i. e. at the maximum value of the fugacity; this is desirable in a model of weighted networks. It is straightforward to see that the asymptotic behaviour of $\mathrm{p}(n)$ is the

same as the asymptotic behaviour of $f_w(n)$. Thus $f_w(n) \sim n^{-b}$ will give a power-law distribution with exponent $-b$. This can be achieved with the hop rate form given in equation (6.33).

Collective species condensation

The case where all the species condense collectively onto the same site is most straightforwardly realised when the attraction between particles depends only on the total number of particles and not explicitly on the number of the same species of particle. This corresponds to setting $u_w(n)$ equal to 1 for all n. Thus particles do not specifically attract particles of their own species; they attract all particles of all species equally. With this choice, $F(z)$ takes the form

$$F(z) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_L=0}^{\infty} f_s(X) z^X \ . \tag{6.37}$$

As the summand in (6.37) depends only on X, the sums can be replaced with a single sum over X and a combinatoric factor that enumerates the number of ways that a sum of L positive integer variables can add up to X

$$F(z) = \sum_{X=0}^{\infty} \binom{L-1+X}{L-1} f_s(X) z^X \ . \tag{6.38}$$

This can be seen formally by introducing an extra sum over X and a delta function to constrain X to be the sum over the number of particles of each species at the site

$$F(z) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_L=0}^{\infty} \sum_{X=0}^{\infty} f_s(X) z^X \delta\left(\sum_{k=1}^{L} n_k - X \right) \ . \tag{6.39}$$

This is then evaluated by using the contour integral representation of the delta function (2.16)

$$F(z) = \sum_{X=0}^{\infty} \sum_{n_1=0}^{\infty} \cdots \sum_{n_L=0}^{\infty} f_s(X) z^X \oint \frac{\mathrm{d}v}{2\pi\iota} \frac{v^{\sum_k n_k}}{v^{X+1}} \ , \tag{6.40}$$

the combinatoric factor then comes from the integral and is found by the theorem of residues [10].

The simplified form of $F(z)$, as given in equation (6.38), gives the density equation

$$\rho = \frac{1}{L} \frac{\sum_{X=0}^{\infty} \binom{L-1+X}{L-1} X f_s(X) z^X}{\sum_{X=0}^{\infty} \binom{L-1+X}{L-1} f_s(X) z^X} \ . \tag{6.41}$$

As before it is the convergence properties of this equation that determine the existence of a condensation transition: if the RHS of (6.41) converges then condensation will occur above some finite density of particles of each species. For the RHS to converge,

the second fraction must converge to a number of $\mathcal{O}(L)$ at the maximum possible value for the fugacity. This is possible for appropriate choices of the hop rate due to the dependence on X in the numerator: the average value of X is expected to be of $\mathcal{O}(L)$ and greater. Expanding the binomial factor for large X yields

$$\binom{L-1+X}{L-1} \simeq \frac{X^{L-1}}{(L-1)!}\,, \tag{6.42}$$

which leads to an approximate version of the density equation that has the same convergence properties.

$$\rho = \frac{1}{L}\frac{\sum_{X=0}^{\infty} X^{L} f_s(X) z^X}{\sum_{X=0}^{\infty} X^{L-1} f_s(X) z^X}\,. \tag{6.43}$$

From this it is straightforward to show that the form $f_s(X) \sim \beta_s^{-X} X^{-b_s}$, with $b_s > 1 + L$ is sufficient for the second fraction in the RHS of (6.43) to converge. However, the overall RHS of equation (6.43) must converge to a finite value for condensation. Thus the second fraction in the RHS must converge to a number $\mathcal{O}(L)$. To find out what the full expression converges to requires the small X behaviour of $f_s(X)$ to be considered. Such an analysis is difficult to perform, but it is found numerically that the form $f_s(X) = [\beta_s(1 + b_s/L)]^{-X}$ for $0 \leq X < L$ and with $b_s > 1 + L$ is sufficient to give condensation at a finite density of each species of particle. A hop rate which satisfies this has the small X behaviour $u_s(X) = \beta_s(1 + b_s/L)$ for $0 \leq X < L$.

Due to the simplified form of $F(z)$, the expressions for the normalisation and the steady state configuration distribution also take on simplified forms

$$\mathcal{Z}(L, z) = \left[\sum_{X=0}^{\infty} \binom{L-1+X}{L-1} f_s(X) z^X\right]^L\,, \tag{6.44}$$

and

$$\mathrm{P}(\{n_{kl}\}) = \frac{1}{\mathcal{Z}(L, z)} \prod_{l=1}^{L} f_s(X_l) z^{X_l}\,. \tag{6.45}$$

Thus the probability of finding a total of X particles at a site is given by

$$\mathrm{p}(X) = \frac{\binom{L-1+X}{L-1} f_s(X) z^X}{\sum_{X=0}^{\infty} \binom{L-1+X}{L-1} f_s(X) z^X}\,. \tag{6.46}$$

This can be used to directly compare the theory to simulation. Again the above equations are only valid when the system is in an uncondensed state. However, as before, equation (6.46) with the maximum value of z will describe the distribution of particles that do not make up the condensate, in the condensed phase.

As it is often desirable to have power-law distributions in networks, equation (6.46) can also be used to determine which hop rates will give a power-law in-strength distribution at criticality, i. e. at the maximum value of the fugacity. Interestingly, in

this case it is not the behaviour for f_s alone that determines the form of the distribution, but its behaviour in concert with the combinatoric factor. Using the asymptotic form of the combinatoric factor (6.42) shows that $\mathrm{P}(X) \sim X^{L-1} f_s(X)$. Thus $f_s(X) \sim X^{-(L+b)}$ will give a power-law distribution of X with exponent $-(1+b)$. This can be achieved by having a hop rate with the behaviour $u_s(X) \sim 1 + L/X + b/X$. To prevent a divergence a different form has to be chosen for small X.

Effect of both strength and weight interactions

The case where both strength and weight interactions are present is somewhat difficult to analyse. An initial analysis, where the normalisation was expressed as a contour integral and approximated by saddle-point methods seemed to indicate that the strength and weight effects should decouple somehow and so each type of condensation should happen due to the relevant type of interaction only; the form of the other interaction should have no effect. However, in simulations (not presented here) this was seen not to be the case and it was clear that this analysis required more care to cope with the subtleties of the system.

As an alternative, the grand canonical approach presented above seems to be a better starting point to analyse the effects of both strength and weight interactions, but the analysis for this case is still generally quite difficult. A sum over X can be introduced into the expression for $F(z)$ (6.27), as in the collective condensation analysis, to give

$$F(z) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_L=0}^{\infty} \sum_{X=0}^{\infty} f_s(X) \delta \left(\sum_{k=1}^{L} n_k - X \right) \prod_{k=1}^{L} f_w(n_k) z^{n_k} \,. \tag{6.47}$$

The difficulty then lies in evaluating the sums: they are strongly coupled when both interactions are present. Evaluation may be possible when one of the interactions is simple. For example, it may be possible to evaluate the sums over $\{n_k\}$ if f_w is some kind of step function. This would allow the study of the effects of introducing a weak interaction in the weights into a system with a strong strength interaction.

For cases where analytical progress is difficult, numerical simulations of the systems can be performed with a simple Monte Carlo algorithm. Results from such simulations are presented in Section 6.7.

6.7 Numerical results

In this section results are presented from simulations of the model that show the two types of condensation predicted by the theory: *independent species condensation*,

where a finite fraction of the particles of each species condenses onto a randomly located site and this site is independent for each species; and *collective species condensation*, where a finite fraction of all the particles in the system condenses onto a single site. In the network context this corresponds to a finite fraction of the total weight from each node being contained in a single link from that node; and a finite fraction of weight pointing from all nodes to a single node, respectively. Results from the case where both weight and strength interactions are present, which is less easily analysed, are also presented.

The model is simulated using a simple Monte Carlo algorithm. At each time step the following update procedure is followed

1. A species of particle, k, and a site of the lattice, l, are chosen at random.
2. If a particle of this species is present then one of these particles will be removed with probability $u(n_{kl}, X_l)\Delta t$, where u is the hop rate (6.10) and Δt is the time interval. The time interval is chosen such that the probability of a hop occurring is always less than or equal to one.
3. If a particle was removed, then a second site m is chosen at random and a particle of species k is added to this site.

The code that simulates the model is presented in Appendix H.

All simulations were run on lattices of size $L = 100$ and with $L = 100$ species of particle, i. e. on networks with 100 nodes, and for $\mathcal{O}(10^7)$ Monte Carlo sweeps. The first half of each run was used to relax to the steady state and the second half to measure distributions of various quantities. Distributions were measured for

- The number of particles of a given species at a site (link weight).
- The total number of particles at a site (node in-strength).
- The number of species present at a site and the number of sites at which a given species is present (in- and out-degree respectively).

Typical configurations of the system were also output; these are presented as 3-D representations of the adjacency matrix and allow the two types of condensation to be identified visually.

6.7.1 Independent species condensation

The system chosen to display the independent condensation of species is defined by the following choice of the hop rate

$$u(n, X) = 1 + \frac{b_w}{n} , \tag{6.48}$$

with $b_w = 4$. Comparing with (6.10), $u_s(X)$ has been set equal to 1 and thus the interaction between species is absent. This hop rate has the correct asymptotic behaviour to show independent species condensation and power-law distributions of the number of particles of a given species at a site, as discussed in Section 6.6.3.

The simulations were run with 175 particles of each species starting from a random initial configuration. The condensation of each species can clearly be seen from the distribution of the number of particles of a given species at a site (black circles, Figure 6.6 (a)), which is the same for all species in this homogeneous system. The peak at around $n = 150$ represents the condensate; it has an area of order $1/L$ and there are L sites, indicating that there is one site for each species that has an occupation of around 150 particles of the relevant species. This represents an appreciable fraction of the number of particles of each species and so can be identified with a condensation, despite the finite size of the system, i. e. it strongly indicates the presence of a condensation in the thermodynamic limit. This distribution also has an apparently power-law background in addition to the condensate peak, as can be seen by comparison with the critical theory line (dashed line, Figure 6.6 (a))—this line is asymptotically a power-law and it matches well with the data shown. This is the behaviour of a single species ZRP as is expected for this system, which is effectively L uncoupled single species ZRPs.

The distribution of the total number of particles at each site (grey crosses, Figure 6.6 (a)) shows that in the absence of any intra-species interaction the condensed sites are randomly distributed. The first peak in the distribution at around $X = 50$ shows that some sites have no condensates on them, only particles from the background distribution of each species. The second, third, ... peaks correspond to sites with condensates of one, two, ... species plus a background from the species that are not condensed on such sites. This is exactly what is seen if the condensates are randomly and independently located, as has been confirmed by generating a distribution of the total number of particles from a random sampling of the measured distribution of particles of a given species. A typical configuration of the system (Figure 6.7) bears this out: each species can be seen to condense onto a single site, but no ordering of locations of the condensates can be discerned.

The distributions of the number of species present at a site and the number of sites at which a given species are present, i. e. the in- and out-degree distributions, are binomial in nature. This can be confirmed by comparing with a binomial distribution generated from the measured probability that there are no particles of a given species at a site. Thus in the network context the connectivity, i. e. the degree distribution, is similar to that of a random graph (which also have binomial degree distributions,

see Chapter 4 Section 4.4.1). In addition to this connectivity structure there is also some structural detail in the weights: a single link from each node contains a finite fraction of the available weight and the in-strength is a random sum of link weights.

Some of the discussion presented for this case is fairly obvious as the system is simply a set of uncoupled single species ZRPs. However, when a similar system that has a weak attraction between the species is simulated the behaviour is almost identical. Thus the discussion above can also be taken as valid for such a system and the analysis presented in Section 6.6.3 may be a starting point for a perturbative analysis of a more complicated system with both inter- and intra-species interaction.

6.7.2 Collective species condensation

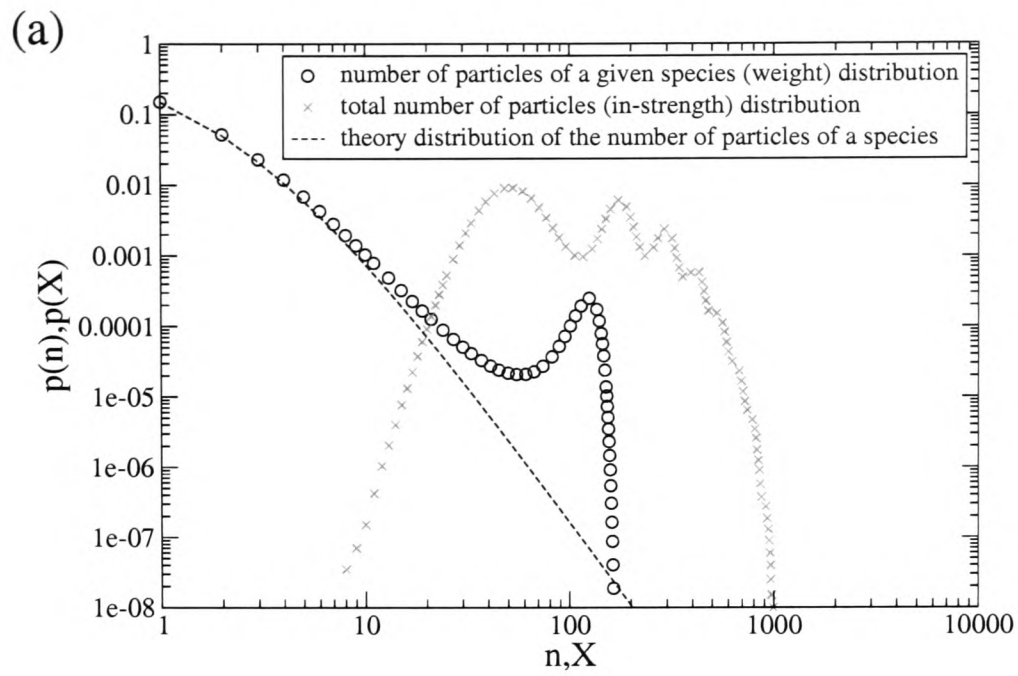
To demonstrate the collective species condensation, the following hop rate was chosen

$$u(n, X) = \begin{cases} 1 + b_s & \text{for} \quad X \leq L \\ 1 + \frac{b_s L}{X} & \text{for} \quad X > L \,, \end{cases} \tag{6.49}$$

with $b_s = 1.05$. Comparing with (6.10), $u_w(n)$ has been set equal to 1 and thus the explicit interaction of particles with their own species has been switched off. This hop rate has the correct low and high end behaviours to show condensation, as discussed in Section 6.6.3 and b_w has been chosen of the form to give a power-law background to the distribution of the total number of particles at a site.

The simulations were run with 1000 particles of each species. The initial configuration for the simulations from which the shown results were taken was all particles at a single site. When starting from random initial conditions the system found itself in a state with several sites with high particle occupancies for the duration of the simulation. Indeed, some simulations started from a configuration with one site having half of all the particles found themselves in this state. This could indicate some instability of the condensate or strong metastability of the state with multiple highly occupied sites. However, when fully formed the condensate appeared to be extremely stable for the duration of the run and similar in nature to condensates from systems with slightly different hop rates (i. e. $b_s > 1 + \mathcal{O}(1/L)$) that were reached very quickly from random initial conditions. Previous knowledge of multi-species ZRPs and the theory also indicate that the condensed phase is the true steady state. Thus it is expected that, given enough time, the system will find itself in the condensed state.

The collective condensation can clearly be seen from the distribution of the total number of particles at a site (grey crosses, Figure 6.8 (a)). The peak at around $X = 85000$ corresponds to the condensate: it has an area of order $1/L$ and there are L sites, indicating that a single site has an occupation of around 85000 particles which is an



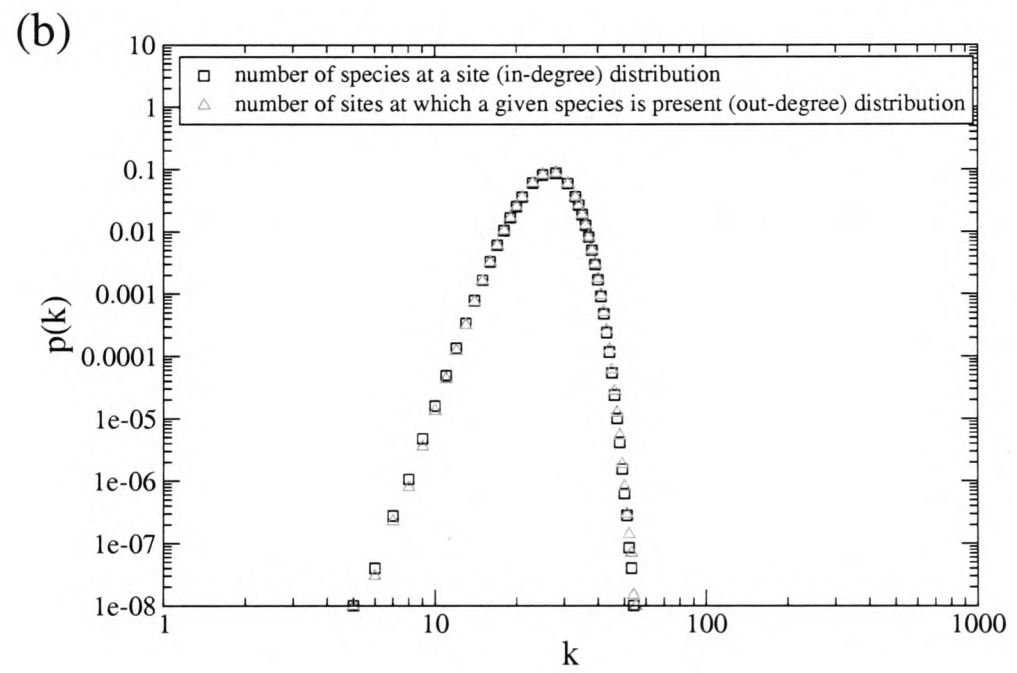


Figure 6.6: Distributions from simulations of the independent species condensation system with hop rate $u(n, X) = u_w(n) = 1 + 4/n$. Simulations were run with a global density of 1.75 of each species of particle, on a lattice with $L = 100$ sites, for $\mathcal{O}(10^7)$ Monte Carlo sweeps and with random initial conditions. (a) Distributions of the number of particles of a given species and the total number of particles at a site, corresponding to the distributions of weight among links and in-strength among nodes in the network context, respectively. Also shown is the expected critical behaviour of the number of particles of a given species at a site, as predicted by the theory (6.36). (b) Distributions of the number of species present at a site and the number of sites at which a given species is present, corresponding to distributions of in- and out-degree in the network context, respectively.

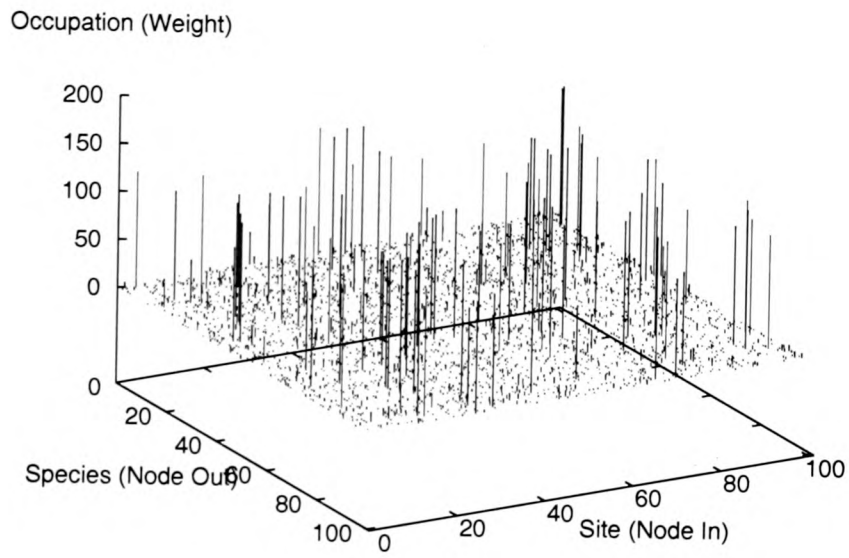


Figure 6.7: Typical configuration of the system output from simulations of the independent species condensation system with hop rate $u(n, X) = u_w(n) = 1 + 4/n$. Simulations were run with a global density of 1.75 of each species of particle on a lattice with $L = 100$ sites for $\mathcal{O}(10^7)$ Monte Carlo sweeps with random initial conditions. Rows represent the species of particle, columns represent sites of the lattice and the height of the bars represents the number of particles.

appreciable fraction of the total number of particles in the system $L \times N = 100 \times 1000$. The other peak at around $X = 100$ corresponds to the sites that do not contain a condensate. After this peak there is a region of apparent power-law behaviour, as predicted by the theory. The theory line generated from equation (6.46) (dashed line, Figure 6.8 (a)) agrees very closely with the background distribution measured from simulations, i. e. it closely reproduces the measured distribution except for the condensate peak.

The distribution of the number of particles of a given species at a site (black circles, Figure 6.8 (b)) also shows condensation behaviour as can be seen in the peak at around $n = 850$. If the interaction between species were to be switched off leaving L uncoupled single-species ZRPs, it is known that the chosen $u_s(n)$ (= constant) would not give condensation in this fully connected homogeneous system—see for example [4]. Thus the attraction that depends only on the total number of particles has induced condensation of the individual species; the full condensate is given by condensates of all the species being located at the same site. The typical configuration of the system shown in Figure 6.9 backs this up: the full condensate is clearly composed of condensates of each of the species all located at the same site. The distribution of the number of particles of a given species at a site also has an apparent power-law piece to it with an exponent that is very close to that seen for the distribution of the total number of particles at a site.

The distribution for the number of sites at which a given species is present (grey triangles, Figure 6.8 (b)) is similar to that for the independent species condensation case: it is a binomial distribution and, as before, matches a binomial distribution constructed from the probability that a site will contain no particles of a given species. Thus the out-degree distribution is like that of a random graph. The distribution of the number of species of particle present at a given site (black squares, Figure 6.8 (b)), however, is significantly different: it has a broader tail and a single data point at $k = L = 100$ with probability $1/L$ representing a single site that contains particles of all species. Thus the in-degree distribution differs from that of a random graph: it has a broader tail meaning there are some nodes with high connectivity with higher probability than in a random graph and there is a single node to which all others point with links that each contain a finite fraction of the weight available to them. The node to which all other nodes point corresponds to the site that contains the condensates of every species of particle. Broadly-tailed degree distributions are often observed in real networks—weighted and unweighted alike. While the tail observed in these simulations is not as broad as many observed, it is at least broader than that of the equivalent random graph. Also, apart from the condensed pieces the in-

strength and weight distributions display power-law tails; this is again something that has been observed in many real networks. Thus, as with the equilibrium networks of Dorogovtsev, Mendes and Samukhin discussed in Chapter 4 Section 4.6.1, certain distributions may take power-law forms at a critical point.

6.7.3 Further behaviour

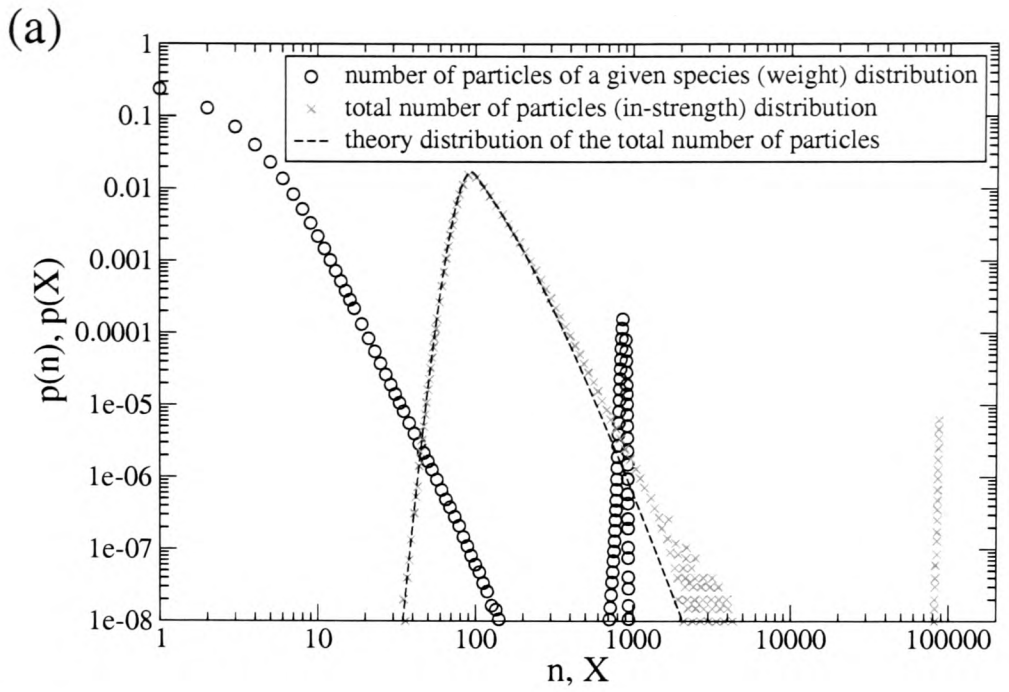
In the preceding sections the somewhat extreme cases of the absence of either intra-species or direct inter-species attraction were considered. This was to allow for a direct comparison with the available theoretical results. Cases where some forms of non-trivial attraction are present in both u_s and u_w, i. e. in both the intra-species and direct inter-species interactions, are difficult to compare directly with theory. However, such systems display interesting behaviour and in this section some results from such a system are presented.

The following form for the hop rate was chosen

$$u(n,X) = \begin{cases} \left(1+\frac{b_w}{n}\right)\left(1+\frac{b_s}{L}\right) & \text{for} \quad X \leq L \\ \left(1+\frac{b_w}{n}\right)\left(1+\frac{b_s}{X}\right) & \text{for} \quad X > L\,, \end{cases} \tag{6.50}$$

with $b_w = 2.5$ and $b_s = 16$. It is known that if the intra-species interaction is switched off for this system (i. e. $u_s(X)$ is set to 1) then the no condensation would occur for this particle number and system size (see for example [4]). If the direct inter-species interaction were to be switched off (i. e. $u_w(n)$ were set to 1) then the theory predicts that no condensation should take place at any finite particle number (see Section 6.6.3). Thus any condensation-like behaviour that is observed is a result of the combination of the two interactions. It was found in simulations that this system does indeed show some condensation-like behaviour.

Simulations were run on a system with 175 particles of each species and from random initial configurations. Starting from a configuration with all particles on one site has no effect on the results. In Figure 6.10 (a), it can be seen that the distribution of the number of particles of a given species at a site (black circles) has a decaying part and a peak at around $n = 150$; this is generally reminiscent of a condensed system. The distribution of the total number of particles at a site (grey crosses, Figure 6.10 (a)) has a peak at $X = 3500$; this is also reminiscent of condensation but the peak is much broader than one usually associated with a clear condensation. The typical configuration data in Figure 6.11 indicate that the peak in the distribution of the total number of particles at a site represents multiple highly occupied sites, each composed of condensates of many, but not all, species.



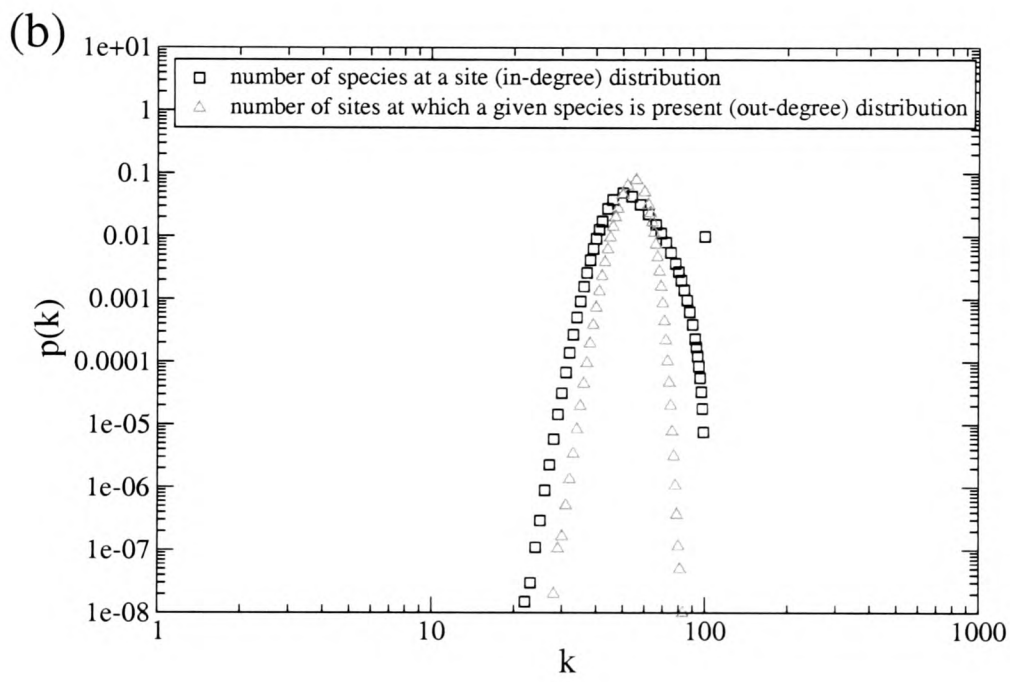


Figure 6.8: Distributions from simulations of the collective species condensation system with hop rate $u(n, X) = u_s(X) = 1 + 1.05$ for $X \leq L$ and $u(n, X) = u_s(X) = 1 + 1.05L/X$ for $X > L$. Simulations were run with a global density of 10 of each species of particle, on a lattice with $L = 100$ sites and for $\mathcal{O}(10^7)$ Monte Carlo sweeps. Due to a long relaxation time for the system, the initial configuration was all particles on one site; this configuration is closer to the expected steady state than a random one. (a) Distributions of the number of particles of a given species and the total number of particles at a site, corresponding to the distributions of weight among links and in-strength among nodes in the network context, respectively. (b) Distributions of the number of species present at a site and the number of sites at which a given species is present, corresponding to distributions of in- and out-degree in the network context, respectively.

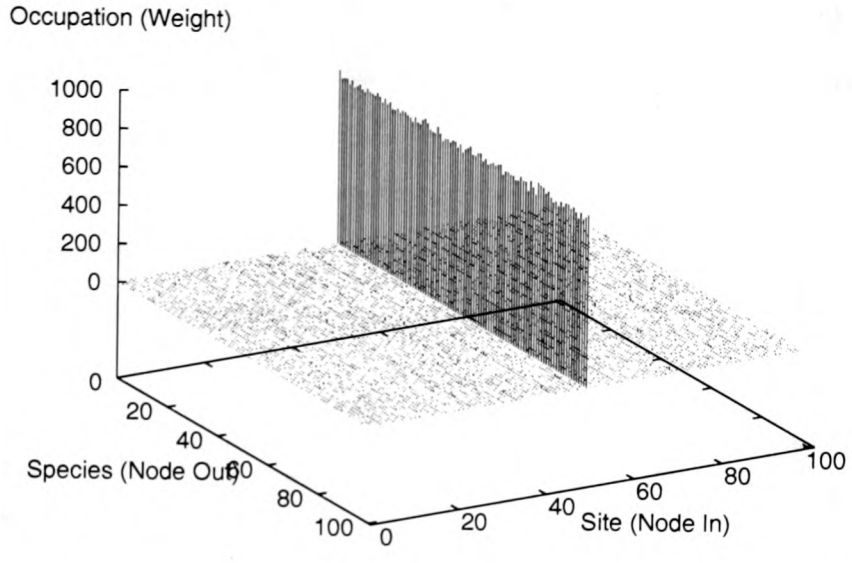


Figure 6.9: Typical configuration of the system output from simulations of the collective species condensation system with hop rate $u(n, X) = u_s(X) = 1 + 1.05$ for $X \leq L$ and $u(n, X) = u_s(X) = 1 + 1.05L/X$ for $X > L$. Simulations were run with a global density of 10 of each species of particle, on a lattice with $L = 100$ sites and for $\mathcal{O}(10^7)$ Monte Carlo sweeps. Due to a long relaxation time for the system, the initial configuration was all particles on one site; this configuration is closer to the expected steady state than a random one. Rows represent the species of particle, columns represent sites of the lattice and the height of the bars represents the number of particles.

Without a direct comparison to theory available it is difficult to interpret these results. They could be a finite size effect that is not seen in larger systems or it could be a true condensation that is being adversely affected by the small size of the system. In either case the results are interesting from a networks point of view as real networks are often of finite size.

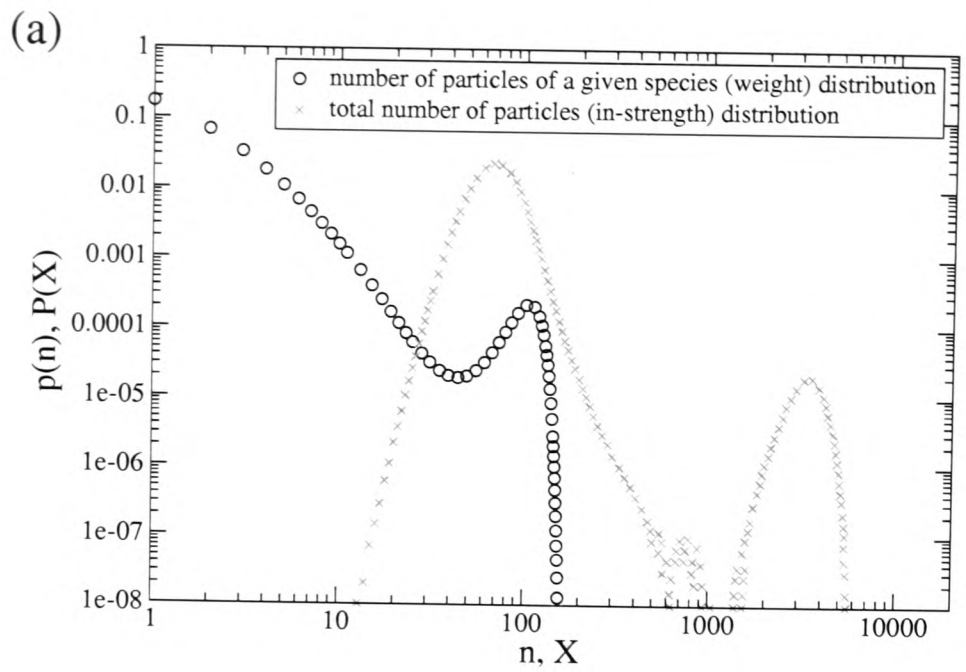
The distributions of the number of species at a site and the number of sites at which a given species are present, i. e. the in- and out-degree distributions, are also interesting for this system. They both have a binomial form for low degree but the in-degree distribution departs from this at high degree: it has a small secondary peak implying that several sites are highly connected but also that a site to which all others point is not certain to be present. Thus, in the networks context, the system gives a network with a connectivity much like that of a random graph except for the existence of several well-connected 'hub' nodes which would not be present in the random case. These hub nodes also have many high weight links pointing to them giving them a large in-strength.

6.8 Generalisations

The basic model introduced in Section 6.6.1 can exhibit interesting behaviour that is relevant to weighted networks, as discussed in the previous two sections. A useful feature of this simple model was that the probability of a configuration in the steady state had a factorised form, making it generally amenable to analysis. A fortunate circumstance is that there are many generalisations of the model that retain the factorised steady state property. Several of these generalisations are discussed briefly in this section.

6.8.1 Disorder and arbitrary site to site transition probabilities

As a first example, it is straightforward to generalise to a case where particles can hop under arbitrary site to site transition probabilities and with rates that depend on the site that is being hopped from—this is similar to the generalisation of the basic ZRP to an arbitrary lattice with disordered sites. In the network context the site dependence of the hop rates allows fitness of the nodes to be modelled, i. e. the inherent ability of a node to retain strength without necessarily having high weight links pointing to it, as discussed for growing models in Section 6.3.4. The case of arbitrary site to site transition probabilities corresponds to a unit of weight pointing to a particular node having prescribed probabilities to rewire to point to other nodes; this could be used to model units of weight preferentially rewiring to nodes that are in some way connected



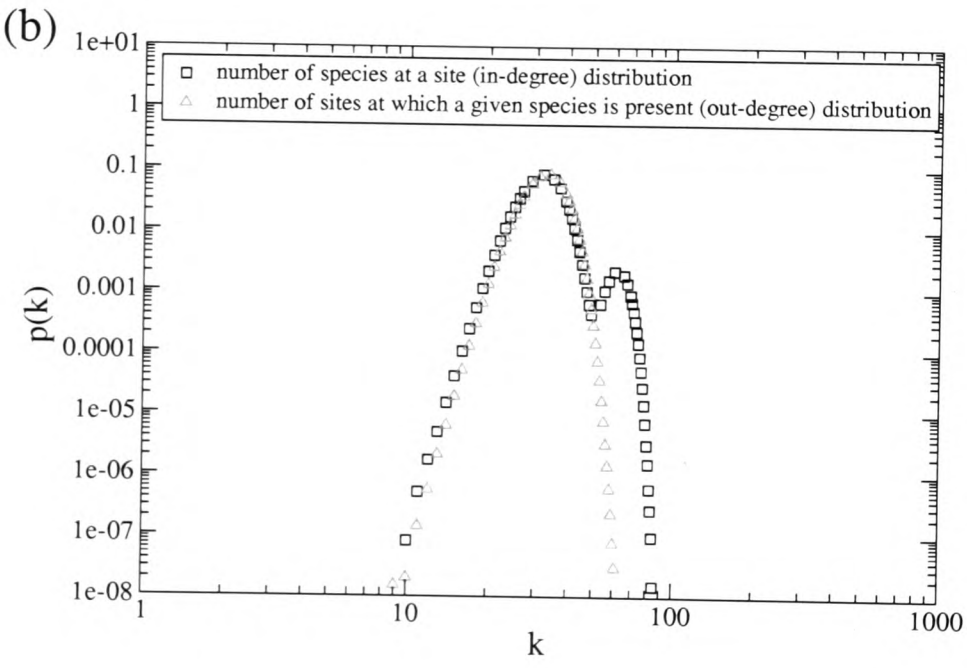


Figure 6.10: Distributions from simulations of the system with both direct intra- and inter-species attractions present with hop rate $u(n, X) = (1 + 2.5/n)(1 + 16/L)$ for $X \leq L$ and $u(n, X) = (1 + 2.5/n)(1 + 16/X)$ for $X > L$. Simulations were run with a global density of 1.75 of each species of particle on a lattice with $L = 100$ sites for $\mathcal{O}(10^7)$ Monte Carlo sweeps from random initial conditions. (a) Distributions of the number of particles of a given species and the total number of particles at a site, corresponding to the distributions of weight among links and in-strength among nodes in the network context, respectively. (b) Distributions of the number of species present at a site and the number of sites at which a given species is present, corresponding to distributions of in- and out-degree in the network context, respectively.

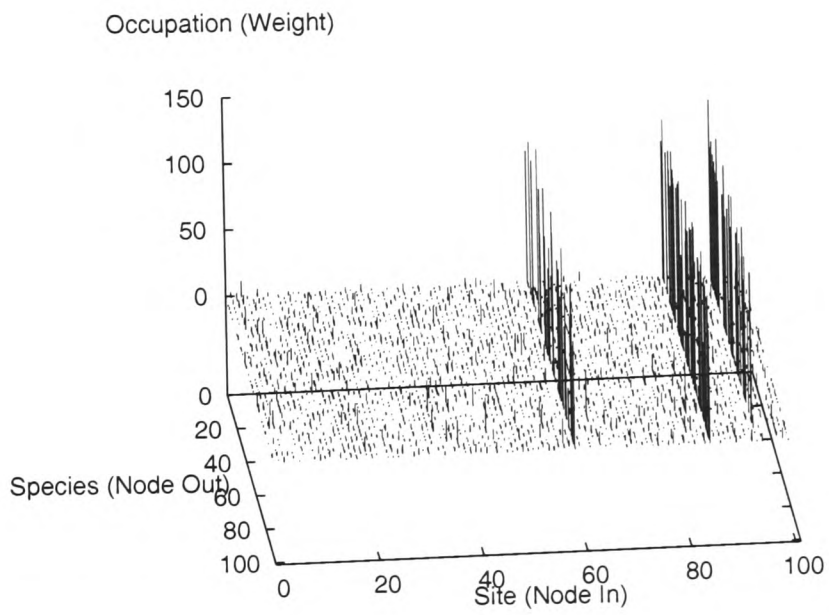


Figure 6.11: Typical configuration of the system output from simulations of the system with both direct intra- and inter-species attractions present with hop rate $u(n, X) = (1 + 2.5/n)(1 + 16/L)$ for $X \leq L$ and $u(n, X) = (1 + 2.5/n)(1 + 16/X)$ for $X > L$. Simulations were run with a global density of 1.75 of each species of particle on a lattice with $L = 100$ sites for $\mathcal{O}(10^7)$ Monte Carlo sweeps from random initial conditions. Rows represent the species of particle, columns represent sites of the lattice and the height of the bars represents the number of particles.

with the node they currently point to for example. Such a system is realised with the following form for the hop rates

$$u_{kl}(n_{kl}, X_l) = u_{w;kl}(n_{kl}) u_{s;l}(X_l) \,, \tag{6.51}$$

where the kl subscripts on u_w denote that the rate now depends on species k and site l and the l subscript on u_s denotes that the rate now depends on the site l. Also required to realise this generalisation is the fact that a particle of species k hopping from site l will hop to site l' with probability

$$\mathrm{W}_k(l', l) \,, \tag{6.52}$$

such that a unique solution can be found for a single particle hopping with rate 1 under these transition probabilities. The steady state probability of a configuration takes the following form for this generalisation

$$\mathrm{P}(\{n_{kl}\}) = \frac{1}{\mathrm{Z}(N, L)} \prod_{l=1}^{L} \left[f_{s;l}(X_l) \prod_{k=1}^{L} \mathrm{r}_{kl} f_{w;kl}(n_{kl}) \right] \,, \tag{6.53}$$

where r_{kl} is the steady state probability of finding a random walker that moves under the site to site transition probabilities $\mathrm{W}_k(l \to l')$ on site l and is defined through

$$\mathrm{r}_{kl} = \sum_{l'} \mathrm{W}_k(l, l') \mathrm{r}_{kl'} \,, \tag{6.54}$$

and where $f_{s;l}(X)$ and $f_{w;kl}(n)$ are given by

$$f_{s;l}(X) = \begin{cases} \displaystyle\prod_{m=1}^{X} \frac{1}{u_{s;l}(m)} & \text{for } X > 0 \\ 1 & \text{for } X = 0 \end{cases} \tag{6.55}$$

$$f_{w;kl}(n) = \begin{cases} \displaystyle\prod_{m=1}^{n} \frac{1}{u_{w;kl}(m)} & \text{for } n > 0 \\ 1 & \text{for } n = 0 \,. \end{cases} \tag{6.56}$$

This steady state can be proven by inserting (6.53) into the balance condition (as the site to site transition probabilities are general at this point, the system may be nonequilibrium and not able to satisfy a *detailed* balance condition)

$$\sum_{k,l,l'} \mathrm{P}(\{n_{kl}\}) u_{w;kl}(n_{kl}) u_{s;l}(X_l) \mathrm{W}_k(l', l) =$$
$$\sum_{k,l,l'} \mathrm{P}(\{\ldots, n_{kl} - 1, \ldots, n_{kl'} + 1, \ldots\}) u_{w;kl'}(n_{kl'} + 1) u_{s;l'}(X_{l'}) \mathrm{W}_k(l, l') \,, \tag{6.57}$$

which, noting that the site to site transition probabilities must be normalised and assuming that terms in the sums over k and l equate, recovers (6.54) and the steady state is seen to hold. Of course, for the steady state solution (6.53) to be unique, the part of the hop rate that is dependent on the total number of particles at the site must be independent of species.

6.8.2 Dependence of hop rates on source and target nodes

Another reasonably straightforward generalisation is to allow the hop rates to depend on the state of both the source and target sites, i. e. a particle of species k hops from site l to site l' with a rate

$$u_{kl}(n_{kl}; X_l)t_{kl'}(n_{kl'}, X_{l'}) \; . \tag{6.58}$$

In the networks context this allows for a preferential attachment rewiring dynamics where units of weight can rewire preferentially to links with high weight and/or nodes with high in-strength. This allows more direct comparisons to be made with network models that grow or rearrange under preferential attachment type rules. This generalisation is closely related to misanthrope processes, as discussed in [4]. The generalisation itself has been discussed in greater detail in [173] where the constraint that the hop rates depend only on the number of particles of the same species and on the total number of particles is also relaxed.

6.8.3 Positive and negative weights

A generalisation to include negative weights is also straightforward. Positive and negative weights have been used to model friendship and animosity in social networks. Such social networks are an example where continual rewiring and changing of the weights of the links is realistic. A model of student relationships in a class which considers positive and negative weights that rewire in single unit steps has been presented in [174].

6.8.4 Continuous masses

The basic ZRP can be generalised to treat continuous masses, where instead of particles there is continuous mass and arbitrary amounts of this mass can move between sites, as has been discussed in [175]. A similar generalisation can be made for the basic model in this chapter, following a similar line to the work in [175]. In the network context this would allow non-integer weights and strengths to be treated.

6.8.5 Variable out-strengths

A particularly straightforward generalisation is to allow a different number of each species of particle, i. e. each site to have a different out-strength. This does not alter the steady state or its derivation, only the analysis of the condensation transitions. It is also possible to generalise to a case where the number of particles of each species can change under the dynamics by allowing particles to 'change their spots' and switch between species. The factorised steady state is retained if particles either hop or change species at any given event, i. e. they do not hop and change species at the same time. The species change can depend on the number of particles of the current species. Thus this generalisation allows for the out-strength to change in much the same way that the in-strength changed in the basic model presented in Section 6.6.1. With the hop rates

$$u(n_{kl}, X_l) = u_w(n_{kl}) u_{is}(X_l) \ , \tag{6.59}$$

for a particle to jump between sites, which depends upon the number of particles of the same species and the total number of particles at a site, and

$$v(n_{kl}, N_k) = v_w(n_{kl}) v_{os}(N_k) \ , \tag{6.60}$$

for a particle to switch species, which depends on the number of particles of the same species at a site and the total number of particles of that species in the system, the steady state takes on the form

$$\mathrm{P}(\{n_{kl}\}) = \frac{1}{\mathrm{Z}(L, \sum_k N_k)} \left[\prod_{l=1}^{L} f_{is}(X_l) \right] \left[\prod_{k=1}^{L} f_{os}(N_k) \right] \left[\prod_{k,l=1}^{L} f_w(n_{kl}) \right] \ . \tag{6.61}$$

The f functions are given by

$$f_{is}(X) = \begin{cases} \displaystyle\prod_{m=1}^{X} \frac{1}{u_{is}(m)} & \text{for } X > 0 \\ 1 & \text{for } X = 0 \end{cases} \tag{6.62}$$

$$f_{os}(N) = \begin{cases} \displaystyle\prod_{m=1}^{N} \frac{1}{u_{os}(m)} & \text{for } N > 0 \\ 1 & = \text{for } N = 0 \end{cases} \tag{6.63}$$

$$f_w(n) = \begin{cases} \displaystyle\prod_{m=1}^{n} \frac{1}{u_w(m)} & \text{for } n > 0 \\ 1 & \text{for } n = 0 \end{cases} \tag{6.64}$$

The *is* and *os* subscripts stand for in-strength and out-strength as the quantities they label depend on the relevant one of these two in the networks context.

This system has yet to be properly analysed and doing so may prove difficult. However, this model is interesting from the networks point of view, as it allows the evolution of both in-strength and out-strength and a model of a network that allows both these to evolve is perhaps more realistic. Thus this generalisation may be worthy of further study. A speculative possibility for the behaviour of this system would be a condensation of weight onto a single link, perhaps with power-law background distributions of the weight, in-strength and out-strength. If this behaviour can be realised by the model, then it would certainly be interesting for the networks viewpoint as it would give an extra quantity, the out-strength, that is distributed according to a power-law at criticality.

6.8.6 Non-conservation

A further generalisation would be to allow some non-conservation in the model, in a similar way to the generalisation of the basic ZRP discussed in Chapter 5. This could again allow the model to have critical, i. e. power-law behaviour, in a region of phase space rather than just at a critical point. Thus the model would give realistic weight and strength behaviour without fine-tuning of parameters.

6.9 Summary

In this chapter a simple model of a directed weighted network based on a multi-species ZRP has been discussed. Particles of the ZRP were considered as units of weight of links in a dynamically rewiring weighted network. The species of a particle was used to encode where the corresponding link pointed from and the site at which the particle was located was used to encode where the link pointed to. Thus each node had a fixed amount of weight pointing from it, but the total weight pointing towards a node or (in-strength) was allowed to vary. The model was found to exhibit two distinct kinds of condensation transition: independent condensation of species of particle and collective condensation of all species of particle, corresponding to condensation of weight onto a single link from each node and condensation of in-strength onto a single node, respectively. At criticality the model was seen to show power-law distributions both for link weight and node in-strength. This is realistic from a networks point of view as power-law weight and strength distributions have been observed in several real networks.

Some of the results of this chapter along with a more in-depth discussion of some of the generalisations have been published in [173].

Chapter 7

Conclusion

The work in this thesis has been mainly concerned with a simple hopping-particle model known as the zero-range process (ZRP) and its behaviour and applications. In the ZRP particles reside on a lattice and hop between sites with rates that depend on the properties of and the number of particles on the site of departure. The ZRP has several advantageous properties including a simple and exactly solvable steady state making it particularly amenable to analysis. Particular attention was paid to condensation transitions in the ZRP, where a finite fraction of the total number of particles condenses onto a single site, and to the application of the ZRP to complex networks.

After a propaedeutic chapter on the basic properties of the ZRP, in Chapter 3 a finite size effect where the current density diagram of a finite system overshoots the thermodynamic predictions was studied in detail for a simple system. The simple system chosen was one with a single defect site at which the hop rate decreased with an increasing number of particles; all other sites had constant hop rates. With this simple system, analysis of the overshoot was possible and the overshoot was found to be a continuation of the fluid phase to higher densities than would be possible in a thermodynamic system along with a region in which the condensate was present, but unstable. This behaviour was compared favourably with that seen in data taken from systems of real traffic and an existing interpretation of the ZRP as a traffic model was discussed. While the single defect site system does not have an obvious meaning in terms of a traffic model (the ZRP traffic model was based around a homogeneous system), it is noteworthy that it displays similar behaviour.

A similar system with two defect sites was also studied and showed that effects due to competition between two defect sites are possible where a condensate initially formed on one defect site and then moved to the other on increasing the density. It

would be interesting to see what would happen in a system with many defect sites and how the behaviour in this system relates to that seen in traffic.

From the analysis of the single defect site system it seemed that the condensation happened via a different mechanism than for the homogeneous system, yet both can produce very similar finite size effects. Thus it would be interesting to see in more detail how the two relate and how, for example, by adding identical defect sites to a system one crosses over to the other.

Following the study of this finite size effect, the application of the ZRP to complex networks was turned to, beginning with a brief review of network theory, behaviour seen in real networks and some of the many models that have been proposed to explain this behaviour in Chapter 4. An interpretation of the ZRP as a network model where particles represent link ends and sites represent nodes was also discussed and was shown to be closely related to some existing models of networks. The ZRP as a network model was able to reproduce a power-law tail to the degree distribution but, as with the related models, only at the critical point of the condensation transition.

Power-law degree distributions seem to be quite generic behaviour for real networks and so a model that reproduces this behaviour only at a critical point is somewhat unsatisfactory. To alleviate this problem, a generalised ZRP with creation and annihilation of particles was introduced in Chapter 5. With appropriate choices for the rates of creation and annihilation the model was found to show critical behaviour in a region of parameter space rather than just at a critical point. The full phase diagram for the model was also found, showing low density, critical and high density phases, with the critical phase having further subdivisions with distinct behaviours. The same principle was then applied to a model of networks by adding creation and annihilation of links to the basic ZRP network model; this had subtle differences to the ZRP model due to the fact that, for a network, annihilating a link corresponds to annihilating a pair of correlated particles. Despite these subtleties, the network model was found to have the same phases as the non-conserving ZRP model.

With a critical phase, the addition of non-conservation to the network model made the power-law degree distribution a more generic phenomenon, thus better representing real networks. However, the dynamics of the model do not appear to have a realistic interpretation for real networks, i. e. the majority of networks do not appear to change their links in the same manner as this model. Despite this, the model still stands as a simple model of a network and it is perhaps worth investigating whether more realistic network models can be found that are based on similar principles. The dynamics of the model also attracts comparison with some ideas from self-organised criticality and it would be interesting to see if further connections between network

models and self-organised criticality can be made. As complex networks apparently organise themselves into complex structures, it is perhaps surprising that many more comparisons do not already exist in the literature.

A further point is that the model only really reproduces the degree distribution realistically: there is nothing present in the model that should generate clustering, degree-degree correlations, etc., other than in trivial senses. However, it is expected that the model will have the small-world property due to its random nature. Thus it would also be interesting to investigate these properties and if, as expected, they are found to be unrealistic, to try to generalise the model to reproduce more realistic behaviour for these and other properties. Even if doing so causes the model to become intractable, the basic model could still be useful as a null model with which real data and future models can be compared. This is in part due to the fact that analytical progress can be made with the model.

A further application of the ZRP to networks was then studied in the form of a novel mapping of a multi-species ZRP to a model of a directed, weighted network in Chapter 6. In this mapping, particles are regarded as units of weight of links, the species of the particle represents the node from which the link is pointing and the site at which the particle is located represents the node to which the link is pointing. Thus there is a species of particle, and a site, for every node in the network. The multi-species ZRP has a steady state with a simple, factorised form that can be solved for hop rates that depend on the number of particles of a given species at a site and the total number of particles at a site, thus particles can interact directly with their own species and also indirectly with all species corresponding to weight and in-strength interactions in the networks context.

The rules of the model allow for two types of condensation transition: independent condensation of species, where each species condenses onto independent, randomly located sites; and collective condensation where each species condenses at the same site. These correspond to condensations of weight and in-strength in the networks context. At criticality the distributions of weight and in-strength can take power-law forms. Power-law weight and strength distributions have been observed in several real systems. Thus, like the basic ZRP network model, the multi-species model of a directed, weighted network shows realistic behaviour for some properties at the critical point and it would be interesting to see if non-conservation could be applied in the same way as to the basic network model to give power-law behaviour under more general conditions.

The multi-species ZRP directed, weighted network also has several other appealing properties. The mapping between the ZRP system and the network system is exact:

the configuration of the particles describes uniquely a configuration of the network. Thus there is no need to reconstruct all possible networks from a given configuration, as for the basic ZRP network model. Also there are many generalisations of the multi-species model that retain the factorised steady state and these can give dynamics of the system that can be considered as more realistic than in the basic ZRP network and non-conserving network cases. For example, with the continuous mass generalisation of the multi-species ZRP, rewirings could be made that take any amount of mass: thus both minor and major rewirings may take place. It is certainly more realistic for many systems that many rewirings would be minor weight rearrangements, with occasional complete and sudden rewirings of entire strong links, as opposed to rapid and frequent rewiring of links as for the basic ZRP network model.

As with the basic ZRP network model, the multi-species model only reproduces some of the behaviour seen in networks, specifically power-law in-strength and weight distributions. The behaviour of the in- and out-degree distributions was not clear for the small systems simulated here. A possible further project would be to simulate much larger systems to see what happens to the degree distribution for these systems. It may also be worthwhile seeing if the degree distributions can be calculated analytically. Something that could be easily measured is topology-weight correlations, although again it would be useful to be able to simulate larger systems to get a clearer picture of this. Also other properties of the system, such as the clustering and node-node strength correlations, could be investigated. In all cases, if the behaviour did not match that seen in real networks it would be interesting to see if the model can be generalised even further to treat these properties.

Overall, the ZRP has been seen to be a model with interesting properties that make it applicable as a model of real systems including traffic systems and, in particular, complex networks. In the context of traffic systems the ZRP has a finite size effect around the condensation transition that matches quite closely behaviour seen in real traffic in the form of an overshoot in the current density diagram. This work includes a detailed study of this effect, something which has not been done for many of the traffic models, and while the particular system the analysis was performed in does not have a direct traffic interpretation, the study can hopefully be extended to systems that do. In the networks context the ZRP found success in reproducing the power-law degree distribution for unweighted networks and, in a multi-species form, success in reproducing power-law weight and in-strength distributions in directed, weighted networks. In both cases this was seen at the critical point of a condensation transition. The advantage of studying such systems in terms of the ZRP is that this is a well-studied model about which much is known. Thus by showing that the ZRP can be

used as a model of the system, many results of the model are known or can be derived straightforwardly and these will be available rapidly for comparison with empirical data. In this sense the ZRP model will serve as a useful null model, even if it only provides qualitatively reasonable behaviour for some aspects of the system. The ZRP is also a simple and quite general system, with many other possibilities to model real systems. As such, anything new discovered when investigating a particular application is in the simple and general language of the ZRP and can be applied to any other systems that the ZRP can be found to model in some way.

The basic ZRP network model is closely related to several existing models and this relation has been clarified, showing that it is possible to gain scale-free behaviour from 'anti-preferential detachment' rules as well as the more common preferential attachment rules. After the development of the BA model, it was speculated that both growth and preferential attachment were necessary ingredients for obtaining a power-law degree distribution in a network [77]. Since then, a lot of evidence controverting this has been put forward. Thus, this study adds to this as the ZRP network model represents a network that produces power-law degree distribution with neither growth or preferential attachment. In particular, this study has shown that, for rewiring networks at least, 'anti-preferential detachment' rules can produce the same steady state distributions as for preferential attachment rules. This is significant, as with 'anti-preferential detachment' as implemented in the ZRP network model, the evolution rates depend only on local information (i. e. the degree of the node from which a rewiring takes place), whereas for preferential attachment, the rates depend on non-local information (i. e. the degree of every node in the network).

The addition of non-conservation to the ZRP network model made it distinct from the existing models and proved that it is possible to modify, using only simple rules, network models that give power-law behaviour only at a critical point—as is the case for several existing models—to give more generic power-law behaviour. It is expected that the application of non-conservation in an appropriate fashion to these existing models will also cause more generic power-law behaviour. The fact that the non-conservation has been studied in the general framework of the ZRP again means that the results from this can be easily applied to any other system that can be modelled by a ZRP. The addition of non-conservation also allowed a comparison with self-organised criticality something that could perhaps be further explored for network models in general.

One interesting thing that should be noted from the non-conserving ZRP network model is the fact that despite possessing a critical phase, in that the degree distribution tends to a power law asymptotically, the degree distribution in this phase can differ

strongly from a pure power law, even for quite large systems. This could indicate that for many real systems the power laws observed are not quite as pure as is believed; they may in fact be power laws with exponential cut-offs or high degree bumps as seen in the critical phase of the non-conserving ZRP network model. It has already been argued that the statistics available for many real networks are not good enough for accurate fitting to power laws, see for example [8], even for the WWW which had $\mathcal{O}(10^8)$ nodes in 1999 and is still growing rapidly.

The mapping of the multi-species ZRP to a model of directed, weighted networks is actually completely different from the mapping of the single-species ZRP to a network and represents a novel mapping of a particle-lattice model to a network model. The fact that a configuration of a directed, weighted network can be mapped exactly onto a configuration of a multi-species particle-lattice model hints that weighted networks—in particular directed, weighted networks—may be a more fruitful area for such mappings than their unweighted brethren. Thus in modelling a directed, weighted network with a particle lattice system, with some success in reproducing realistic behaviour, this work has hopefully paved the way for further such models to be applied to the field of networks. The specifics of the multi-species model also brought to light a new condensation transition where particles of all species condense onto a single site. In the network context, this corresponds to in-strength condensation with induced weight condensations and can yield power-law distributions for both these quantities at criticality.

It would also be generally interesting to see how much further both of the ZRP models could be taken in representing networks realistically, i. e. to see if there are further generalisations that can be made to treat other network properties such as clustering and degree-degree correlations. In this respect the exactly solvable nature of the ZRP and many of its generalisations is a real advantage, as with an exact solution and existing knowledge base the ZRP is a more obvious starting point for further generalisation than some of the existing network models. Even in cases where it is known or expected that a generalisation of the ZRP will no longer have an exact solution, the solution of the basic ZRP could be used as a starting point for a perturbative analysis. Thus the success of the ZRP as a model of networks, and other systems, can hopefully be continued with further generalisations and perhaps give further insight into the organising principles behind the complex behaviour seen in many real networks.

Appendix A

Proof of the Requirements on the Hop Rates for Condensation in the Homogeneous Zero-Range Process

For the homogeneous ZRP the requirement on the hop rates for condensation is that they tend to some finite value β more slowly than $\beta(1+2/n)$. The overall requirement for condensation in the homogeneous system is the convergence of the RHS of (2.14) at the maximum value of the fugacity, $z_{\max}$, which takes the form

$$z_{\max}\frac{F'(z_{\max})}{F(z_{\max})}\,. \tag{A.1}$$

It is straightforward to show that $F(z)$ and $F'(z)$ are monotonically increasing functions of z. From the form of $F(z)$ and the fact that the convergence of a sum is determined by the last infinity of terms it is straightforward to see that if $F'(z)$ converges then $F(z)$ must also converge and thus (A.1) must converge. Therefore the condition for condensation is that $F'(z_{\max})$ is convergent.

The requirement for the convergence of $F'(z_{\max})$ can be found as follows. First the hop rate is assumed to decay to β as $u(n) \sim \beta(1+\xi(n))$, where $\xi(n)$ is some small, decreasing function. Note that with this hop rate, $z_{\max} = \beta$. This form of the hop rate gives

$$f(n) \sim \beta^{-n} \exp\left[-\sum_{m=1}^{n} \ln(1+\xi(n)\right]\,, \tag{A.2}$$

from (2.3). Approximating the sum by an integral and expanding the logarithm for

large n yields

$$f(n) \sim \beta^{-n} \exp\left[-\int \mathrm{d}m\, \xi(n)\right] . \tag{A.3}$$

Inserting this and $z_{\max}$ into the expression for $F'(z)$ (found by differentiating (2.11)) gives

$$F'(z_{\max}) \sim \sum_{n=1}^{\infty} n \exp\left[-\int \mathrm{d}m\, \xi(n)\right] . \tag{A.4}$$

By comparing this with the convergence of the Riemann zeta function, $\int \mathrm{d}m\, \xi(m)$ must be greater than $2\ln(n)$ for convergence. Thus $\xi(m)$ must be greater than $2/n$ and for convergence the hop rate must therefore decay more slowly than $\beta(1 + 2/n)$.

Appendix B

Asymptotic Expansion of the Normalisation Constant of the SDS System to give the Particle Current

The sum (3.36) can be rewritten as

$$\begin{aligned} \mathrm{Z}(N,L) &= \sum_{n=0}^{N} p^{-n} \frac{\Gamma(n+1)\Gamma(b+1)}{\Gamma(n+b+1)} \binom{N+L-n-2}{L-2} \\ &\simeq \int_0^N \mathrm{d}n\, p^{-n} \frac{\Gamma(n+1)\Gamma(b+1)}{\Gamma(n+b+1)} \binom{N+L-n-2}{L-2} . \end{aligned} \tag{B.1}$$

To calculate the asymptotic expansion of this, it must be put into a general form for which the formula for the first order asymptotic expansion exists [18]

$$\begin{aligned} \int \mathrm{d}n\, \mathrm{A}(n) \exp(L\mathrm{B}(n)) \;\simeq\; & \exp\left(L\mathrm{B}\right) \left(\frac{2\pi}{\left|\mathrm{B}^{(2)}\right| L}\right)^{\frac{1}{2}} \\ & \times \left[\mathrm{A} + \frac{1}{L}\left\{\frac{\mathrm{A}^{(2)}}{2\mathrm{B}^{(2)}} + \frac{\mathrm{A}\mathrm{B}^{(4)}}{8\left(\mathrm{B}^{(2)}\right)^2} + \frac{\mathrm{A}^{(1)}\mathrm{B}^{(3)}}{2\left(\mathrm{B}^{(2)}\right)^2} + \frac{5\mathrm{A}\left(\mathrm{B}^{(3)}\right)^2}{24\left(\mathrm{B}^{(2)}\right)^3}\right\} + \mathcal{O}\left(\frac{1}{L^2}\right)\right]\Bigg|_{n=n_0} , \end{aligned} \tag{B.2}$$

where n_0 is the location of the maximum of the exponent function $\mathrm{B}(n)$ and $\gamma^{(m)}(n)$ denotes the m^{th} derivative of $\gamma(n)$ with respect to n. To put equation (B.1) in this form, one first needs to find the expansions of the Gamma functions and the binomial coefficient. The expansion of the relevant Gamma functions for large n, as can be

taken from Stirling's series, is

$$\frac{\Gamma(n+1)}{\Gamma(n+b+1)} \simeq n^{-b}\left[1-\frac{b(b+1)}{2n}\right] \tag{B.3}$$

$$= L^{-b}y^{-b}\left[1-\frac{b(b+1)}{2Ly}\right], \tag{B.4}$$

to first order in L, with $y = n/L$. The first order asymptotic expansion of the binomial coefficient is more unwieldy. However, as the primary interest lies in the first order expansion for the particle current, v, the binomial coefficient needs only to be expanded to zeroth order, as indeed does the Gamma function expansion. To see this, note how the expansion of the binomial coefficient is performed. The binomial coefficient can be represented as an integral in the complex plane as follows

$$\begin{aligned}\binom{N+L-n-2}{L-2} &= \oint \frac{\mathrm{d}w}{2\pi\imath}\frac{(1+w)^{N+L-n-2}}{w^{L-1}} \\ &= \oint \frac{\mathrm{d}w}{2\pi\imath}\frac{w}{(1+w)^2}\exp\left\{L\left[(\rho+1-y)\ln(1+w)-\ln(w)\right]\right\}.\end{aligned} \tag{B.5}$$

This can then be asymptotically expanded in the system size L via the saddle point method as it has been put in the proper form [18].

$$\begin{aligned}\oint \mathrm{d}w\,\alpha(w)\exp\left(L\beta(w)\right) \simeq \exp\left(L\beta\right)\left(\frac{2\pi}{\left|\beta^{(2)}\right|L}\right)^{\frac{1}{2}} e^{\imath\phi} \\ \times\left[\alpha+\frac{1}{L}\left\{\frac{\alpha^{(2)}e^{2\imath\theta}}{2\beta^{(2)}}+\frac{\alpha\beta^{(4)}e^{4\imath\theta}}{8\left(\beta^{(2)}\right)^2}+\frac{\alpha^{(1)}\beta^{(3)}e^{4\imath\theta}}{2\left(\beta^{(2)}\right)^2}+\frac{5\alpha\left(\beta^{(3)}\right)^2e^{6\imath\theta}}{24\left(\beta^{(2)}\right)^3}\right\}\right. \\ \left.+\mathcal{O}\left(\frac{1}{L^2}\right)\right]\Bigg|_{w=w_0},\end{aligned} \tag{B.6}$$

where w_0 is the location of the saddle point of $\beta(w)$, $\gamma^{(n)}$ denotes the n^{th} derivative of γ with respect to w, θ is chosen such that $|\beta^{(2)}(w_0)|e^{\imath\theta} = \beta^2(w_0)$, and $\phi = (\pi-\theta)/2$. Now, when the ratio $\mathrm{Z}(N-1,L)/\mathrm{Z}(N,L)$ is taken to find the particle current, only terms which depend on N explicitly will matter, all other terms will cancel to first order. Furthermore, with terms that depend explicitly on the particle density, $\rho = N/L$, taking the ratio will generate only higher order corrections than the order of the term. To clarify this a simple example is presented. Consider a partition function which had the asymptotic form $\mathrm{Z}(N,L) = C(\rho)+(1/L)D(\rho)$, thus taking the ratio will give

$$\begin{aligned}\frac{C(\rho-1/L)+(1/L)D(\rho-1/L)}{C(\rho)+(1/L)D(\rho)} &= C(\rho-1/L)-C(\rho)+\frac{1}{L}\left\{D(\rho-1/L)-D(\rho)\right\} \\ &\quad+\cdots \\ &= -(1/L)C'(\rho)-(1/L^2)D(\rho)+\cdots,\end{aligned} \tag{B.7}$$

where in the second line Taylor expansions about ρ have been made. Thus it is clear that only the zeroth order expansion in L of a normalisation with the above general form is required to generate the first order expansion of the particle current. This is because the saddle point location w_0 for the binomial coefficient has no N dependence so the only N dependence comes from the $\beta(w)$ and its derivatives, as this depends on ρ. As it only depends on ρ, only the zeroth order is needed to calculate the first order for the particle current.

Taking the ratio of the asymptotic expansions of the two normalisations gives the asymptotic expansion of the particle current

$$v = p\left[1 + \frac{1}{L}\frac{b}{\rho - \rho_c} + \mathcal{O}\left(\frac{1}{L^2}\right)\right] . \tag{B.8}$$

Appendix C

Code to Simulate the Homogeneous Zero-Range Process

```
//-----------------------------------------------------------------
// C++ code to simulate the homogeneous, 1d, totally assymetric
// zero range process
//-----------------------------------------------------------------
// Written by Andrew Angel
// November 2002
// Last Update 19th November 2002
// Compiler: g++
//-----------------------------------------------------------------

#include <iostream.h>
#include <stdlib.h>
#include <stdio.h>
#include <fstream.h>
#include <math.h>
#include <string.h>
#include <strstream.h>
#include "rng.h"

// global constants
```

```
const long double LAMBDA = 3.2;      //parameters for run
const unsigned int MAX_PARTICLES = 2000;
const unsigned int LATTICE_SIZE = 1000;
const unsigned int NO_MC_STEPS = 1000000;
const unsigned int AVG_NO = 250000;
// how many times to avg when in stdy st.
const int SEED = 0; // random num will be seeded from the system time
// if 0
const unsigned int RANDOM_FILL = 0; // 0 -> start all particles on
// one site, random ICs otherwise
const unsigned int FILLED_SITE = 500;  // where the filled site is.
const unsigned int PARTICLE_INTERVAL = 10; // increment for particles


// procedures and functions

void random_particle_fill(unsigned int[LATTICE_SIZE], unsigned int,
                          unsigned int);

void mc_update(unsigned int[LATTICE_SIZE], unsigned int,
               long double[MAX_PARTICLES+1]);

void vel_mc_update(unsigned int[LATTICE_SIZE], unsigned int,
                   long double[MAX_PARTICLES+1],
                   unsigned int[LATTICE_SIZE]);

long double calculate_variance(unsigned int[LATTICE_SIZE],
                               unsigned int, unsigned int);

unsigned int find_most_occupied(unsigned int[LATTICE_SIZE],
                                unsigned int);

void create_look_up(long double[MAX_PARTICLES+1],unsigned int,
                    unsigned int);

long double hop_rate(int x)
```

```
{
  long double val;

  if (x != 0){
    val = (long double) 1.0 + ( LAMBDA / (long double) x );
    val = val / (long double) ( 1.0 + LAMBDA);
  }
  else{
    val = (long double) 0.0;
  }
  return val;
}

// begin main
main()
{
  // declare local variables
  unsigned int config[LATTICE_SIZE] = {0};
// holds the sytems configuration
unsigned int totalSteps = LATTICE_SIZE * NO_MC_STEPS;
unsigned int updatesPerStep = LATTICE_SIZE;  // for readability
long double particleDensity = 0.0;
long double dataVariance = 0.0;
long double avgVariance = 0.0;
unsigned int halfwayStep = NO_MC_STEPS / 2;
// want integer division here
unsigned int outputInterval = halfwayStep / AVG_NO;
// " " " "
long double hopRateLookup[MAX_PARTICLES + 1] = {0};
unsigned int mos;          // most occupied site
long double dataFracMostOcc = 0;   // running total of fmo
long double avgFracMostOcc = 0;    // average of fmo
unsigned int velArray[LATTICE_SIZE] = {0};
// total no of hops made in ss
long double avgVel[LATTICE_SIZE] = {0.0};
// avgd hr for each site
long double overallAvgVel = 0.0;  // hoprate avgd over all sites
```

```
unsigned int i,j,k,l;   // counter variables for loops
unsigned int totalParticles;
char flNmStr[25];

// open data streams
ofstream fVar;
ofstream fFMO;
ofstream fVel;
ofstream fAvgVel;

srand48(time(NULL));  // seed the random number generator

// create the look_up_table
create_look_up(hopRateLookup, LATTICE_SIZE, MAX_PARTICLES);

// name and open files
ostrstream name1(flNmStr, sizeof flNmStr);
name1 << "variance.data" << ends;
fVar.open(flNmStr);
ostrstream name2(flNmStr, sizeof flNmStr);
name2 << "fmo.data" << ends;
fFMO.open(flNmStr);
ostrstream name3(flNmStr, sizeof flNmStr);
name3 << "velocity.data" << ends;
fVel.open(flNmStr);
ostrstream name4(flNmStr, sizeof flNmStr);
name4 << "avgvel.data" << ends;
fAvgVel.open(flNmStr);

for (i=0; i < MAX_PARTICLES; i += PARTICLE_INTERVAL){
// loop over densities
  cout << "Got to " << i+1 << " particles." << endl;

  particleDensity = (long double) (i+1) /
    (long double) LATTICE_SIZE;

  if (RANDOM_FILL == 0){  //random fill or all on one
```

```
  for (j=0 ; j < LATTICE_SIZE; j++){
    config[j] = 0;
  }
  config[FILLED_SITE] = i+1 ;
}
else{
  random_particle_fill(config,(unsigned) i+1,LATTICE_SIZE);
}

for(j=0;j<LATTICE_SIZE;j++){
  velArray[j] = 0;
  avgVel[j] = 0;
}

dataVariance = 0.0;
avgVariance = 0.0;
dataFracMostOcc = 0.0;
avgFracMostOcc = 0.0;
overallAvgVel = 0.0;

for(j=0;j<halfwayStep;j++){
  for(k=0;k<updatesPerStep;k++){
    mc_update(config,LATTICE_SIZE,hopRateLookup);
  }
}

for(j=0;j<AVG_NO;j++){
  dataVariance = dataVariance +
                 calculate_variance(config,LATTICE_SIZE,
                                                    i+1);
  mos = find_most_occupied(config,LATTICE_SIZE);
  dataFracMostOcc = dataFracMostOcc + ( (long double)
                                        config[mos] /
                                        (long double) (i+1) );
  for(k=0;k<outputInterval;k++){
    for(l=0;l<updatesPerStep;l++){
      vel_mc_update(config,LATTICE_SIZE,hopRateLookup,velArray);
```

```
    }
  }
}

avgVariance = dataVariance / (long double) AVG_NO;
fVar << particleDensity << " " << avgVariance << endl;
avgFracMostOcc = dataFracMostOcc / (long double) AVG_NO;
fFMO << particleDensity << " " << avgFracMostOcc << endl;

for(j=0;j<LATTICE_SIZE;j++){
  avgVel[j] = (long double) velArray[j] /
    ( (long double) (AVG_NO * outputInterval * updatesPerStep) );
  fVel << j << " " << particleDensity << " " << avgVel[j] << endl;
  overallAvgVel = overallAvgVel + avgVel[j];
}
fAvgVel << particleDensity << " " << overallAvgVel << endl;
/*  note that the overall average velocity is actually
    LATTICE_SIZE times what it should be, this scales
    the data so it can all be placed on one graph, or
    can define average hopping rate out of site as hops
    from site / hops attempted from site.

    Also note that slightly confusingly the average velocity
    array is output to the velocity data file and not the
    average velocity data file */
}

fVar.close();
fFMO.close();
fVel.close();
fAvgVel.close();

// output stuff to a log file for error checking

ofstream fLog;
ostrstream name5(flNmStr, sizeof flNmStr);
name5 << "output.log" << ends;
```

```
fLog.open(flNmStr);

fLog << "Lambda = " << LAMBDA << endl;
fLog << "Lattice Size = " << LATTICE_SIZE << endl;
fLog << "Maximum Number of Particles = " << MAX_PARTICLES << endl;
fLog << "Number of Monte Carlo Steps = " << NO_MC_STEPS << endl;
fLog << "Number of times averages were taken = " << AVG_NO << endl;
fLog << "Particle Increment = " << PARTICLE_INTERVAL << endl;
fLog << "Seed = " << SEED << endl;
fLog << "Random Fill status was: " << RANDOM_FILL << endl;
fLog << "Filled Site = " << FILLED_SITE << endl;
totalParticles = 0;
for(i=0;i<LATTICE_SIZE;i++){
  totalParticles = totalParticles + config[i];
}
fLog << "Totalled particles at end = " << totalParticles << endl;
fLog.close();
  exit(EXIT_SUCCESS);
}


// procedures and functions in full if not already so above

void create_look_up(long double hopRateLookup[],
                    unsigned int LATTICE_SIZE,
                    unsigned int MAX_PARTICLES)
{
  int i = 0;

  for ( i = 0; i <= MAX_PARTICLES ; i++){
    hopRateLookup[i] = hop_rate(i);
  }
}

void random_particle_fill(unsigned int config[],
                          unsigned int particleNumber,
```

```
                        unsigned int LATTICE_SIZE)
{
  unsigned int i;
  unsigned int randSite;

  for(i=0;i<LATTICE_SIZE;i++){
    config[i] = 0;    // make sure array empty before filling it
  }

  for(i=0;i<particleNumber;i++){
    randSite = (unsigned int)
                  floor(drand48() * (long double) LATTICE_SIZE);
    config[randSite]++;
  }

  return;
}

void mc_update(unsigned int config[], unsigned int LATTICE_SIZE,
               long double hopRateLookup[])
{
  unsigned int site;
  long double diceRoll;

  site = (unsigned int) floor( drand48() * LATTICE_SIZE );
  diceRoll = drand48();

  if (diceRoll < hopRateLookup[config[site]]){
    if(site != 0){
      config[site] = config[site] - 1;
      config[site-1] = config[site-1] + 1;
    }
    else{
      config[site] = config[site] - 1;
      config[LATTICE_SIZE - 1] = config[LATTICE_SIZE - 1] + 1;
    }
  }
```

```
}


void vel_mc_update(unsigned int config[], unsigned int LATTICE_SIZE,
                   long double hopRateLookup[],
                   unsigned int velArray[])
{
  unsigned int site;
  long double diceRoll;

  site = (unsigned int) floor( drand48() * LATTICE_SIZE );
  diceRoll = drand48();

  if (diceRoll < hopRateLookup[config[site]]){
    if(site != 0){
      config[site] = config[site] - 1;
      config[site-1] = config[site-1] + 1;
    }
    else{
      config[site] = config[site] - 1;
      config[LATTICE_SIZE - 1] = config[LATTICE_SIZE - 1] + 1;
    }
    velArray[site] = velArray[site] + 1; //if hop then running total
  }
}

long double calculate_variance(unsigned int config[],
                               unsigned int LATTICE_SIZE,
                               unsigned int particleNumber)
{
  unsigned int i;
  unsigned int sumOccSquared;
  long double meanOcc, meanOccSquared, meanSquaredOcc, val;

  i=0;
  sumOccSquared=0;
  meanOcc = 0.0;
```

```
  meanOccSquared = 0.0;
  meanSquaredOcc = 0.0;
  val = 0.0;

  for(i=0;i<LATTICE_SIZE;i++){
    sumOccSquared = sumOccSquared + (config[i]*config[i]);
  }

  meanOcc = (long double) particleNumber
                                / (long double) LATTICE_SIZE;
  meanOccSquared = (meanOcc * meanOcc);
  meanSquaredOcc = (long double) sumOccSquared
                                    / (long double) LATTICE_SIZE;

  val = (meanSquaredOcc - meanOccSquared)
                           / (long double) LATTICE_SIZE;

  /* variance = mean square occupancy - mean occupancy squared
     have normalised by the LATTICE SIZE to make any
     transition clearer i.e. to convert the variance into an
     order parameter */
    return val;

}

unsigned int find_most_occupied(unsigned int config[],
                                  unsigned int LATTICE_SIZE)
{
  unsigned int i,m,n;
  unsigned int val;

  n = config[0];
  m = 0;
  for(i=1;i<LATTICE_SIZE;i++){
    if(config[i] > n){
      n = config[i];
      m = i;
```

```
    }
  }
  val = m;
  return val;
}
```

Appendix D

Code to Calculate the Normalisation Constant of the Homogeneous Zero-Range Process Exactly

```
//-------------------------------------------------------------
// C++ code to calculate -ln(Z(M,L)) from the homogeneous
// zero range process in one dimension.  From this the
// program also calculates the velocity, first and second
// derivatives in the free energy, and the probability
// distribution for the occupancy of any site.
//-------------------------------------------------------------
// Written by Andrew Angel
// December 2002
// Last Update Apr 29 2003
// Compiler: g++
//-------------------------------------------------------------

#include <iostream.h>
#include <stdlib.h>
#include <fstream.h>
#include <math.h>
#include <string.h>
```

```
#include <strstream.h>

// global constants
const long double LAMBDA = 2.2;
const int MAX_PARTICLES = 8000;
const int MAX_SITES = 1000;
const int P_S_FACTOR = 10;
const int L_SAMPLES = 100;

// procedures and functions
void partition(int,int);
long double hop_rate(int);

// global variables
long double lnPartitionArray[2][MAX_PARTICLES+2];
long double marginalArray[MAX_PARTICLES+2];

// global pointers
long double *hopRateArray;

// begin main
main()
{
  char flNmStr[35];

  int i,j,p,q,s,t;

  long double particleDensity;
  long double maxDensity;
  long double freeEnergy;
  long double freeEnergyDensity;    // free energy per site
  long double velocity;
  long double firstDiff;
  long double secondDiff;
  long double prob;
  int numberOfSites;
  int particleIncrement;
```

```
int probParticles;

ofstream fFree;
ofstream fFreeDens;
ofstream fMarginal;
ofstream fVelocity;
ofstream fFirstDiff;
ofstream fSecondDiff;
ofstream fProb;
ofstream fLnLnProb;
ofstream fOutput;

hopRateArray = new long double [MAX_PARTICLES+2];

for(i=0;i<=1;i++){
  for(j=0;j<=MAX_PARTICLES+1;j++){
    lnPartitionArray[i][j] = 0.0;
    marginalArray[j]=0.0;
    hopRateArray[j]=0.0;
  }
}

// calculate the hoprates and marginals
marginalArray[0] = 1.0;   // IC
for(i=1;i<=MAX_PARTICLES+1;i++){
  hopRateArray[i] = hop_rate(i);
}
fMarginal.open("marginal.dat");
for(i=1;i<=MAX_PARTICLES+1;i++){
  marginalArray[i] = marginalArray[i-1] / hopRateArray[i];
  fMarginal << i << " " << marginalArray[i] << endl;
  fMarginal << i << " " << hopRateArray[i] << endl;
}
fMarginal.close();

//IC Z(1,L) = f(L)
for(i=0;i<=MAX_PARTICLES+1;i++){
```

```
  lnPartitionArray[1][i] = log(marginalArray[i]);
}

// what interval of sites to generate output for
numberOfSites = (int) floor( (long double)MAX_SITES /
                             (long double)P_S_FACTOR );
maxDensity = (long double) MAX_PARTICLES
                               / (long double) MAX_SITES;

// loop over each site increment call to the partion
// sum calculation function and generate outputs
for(p = numberOfSites; p <= MAX_SITES; p += numberOfSites){

  ostrstream name1(flNmStr,sizeof flNmStr);
  name1 << "F_M=" << p << "_dens.dat" << ends;
  fFree.open(flNmStr);
  ostrstream name2(flNmStr,sizeof flNmStr);
  name2 << "f_M=" << p << "_dens.dat" << ends;
  fFreeDens.open(flNmStr);
  ostrstream name3(flNmStr,sizeof flNmStr);
  name3 << "vel_M=" << p << "_dens.dat" << ends;
  fVelocity.open(flNmStr);
  ostrstream name4(flNmStr,sizeof flNmStr);
  name4 << "der1_M=" << p << "_dens.dat" << ends;
  fFirstDiff.open(flNmStr);
  ostrstream name5(flNmStr,sizeof flNmStr);
  name5 << "der2_M=" << p << "_dens.dat" << ends;
  fSecondDiff.open(flNmStr);

  particleIncrement = (int) floor( (long double)p
                                   / (long double)L_SAMPLES );
  if(particleIncrement<1){
    cout << "Particle Increment is too small, exiting." << endl;
    exit(EXIT_FAILURE);
  }

  partition(p - numberOfSites+1,p);
```

```
for(q=1; q<= maxDensity * p; q+=particleIncrement){
  particleDensity = (long double)q / (long double)p;
  freeEnergy = -lnPartitionArray[p%2][q];
  fFree << particleDensity << " " << freeEnergy << endl;
  freeEnergyDensity = freeEnergy/(long double)p;
  fFreeDens << particleDensity << " "
            << freeEnergyDensity << endl;
  velocity = exp(lnPartitionArray[p%2][q-1]
                              - lnPartitionArray[p%2][q]);
  fVelocity << particleDensity << " " << velocity << endl;
  firstDiff = (lnPartitionArray[p%2][q-1]
               - lnPartitionArray[p%2][q+1])/2.0;
  fFirstDiff << particleDensity << " " << firstDiff << endl;
  secondDiff = p*(2*lnPartitionArray[p%2][q]
                  - lnPartitionArray[p%2][q+1]
                  - lnPartitionArray[p%2][q-1]);
  fSecondDiff << particleDensity << " " << secondDiff << endl;
  }

  cout << "Done " << p << " sites." << endl;

  fFree.close();
  fFreeDens.close();
  fVelocity.close();
  fFirstDiff.close();
  fSecondDiff.close();
}

/* Calculate P(k) the occupancy probability distribution, this
   is the same for all sites in the homogeneous system, also
   calculate just at the end for 1000 sites */

for(s=1;s<=10;s++){
  probParticles = (int)floor(s*((double)(MAX_PARTICLES)
                                /(double)(MAX_SITES))
                                    / (10.0) * MAX_SITES);
```

```
ostrstream name6(flNmStr,sizeof flNmStr);
name6 << "prob_M=" << MAX_SITES << "L=" << (int) probParticles <<
  ".dat" << ends;
fProb.open(flNmStr);
ostrstream name7(flNmStr,sizeof flNmStr);
name7 << "lnpr_M=" << MAX_SITES << "L=" << (int) probParticles <<
  ".dat" << ends;
fLnLnProb.open(flNmStr);
for(t=0;t<=probParticles;t++){
  prob = marginalArray[t] * exp(lnPartitionArray[(MAX_SITES-1)%2]
                                [probParticles-t] -
                                lnPartitionArray[MAX_SITES%2]
                                [probParticles]);
  fProb << t << " " << prob << endl;
  if(t!=0){
    fLnLnProb << log(t) << " " << log(prob) << endl;
  }
}
fProb.close();
  fLnLnProb.close();
}
cout << "Calculated probabilities." << endl;

fOutput.open("output.log");
fOutput << "LAMBDA = " << LAMBDA << endl;
fOutput << "MAX_PARTICLES = " << MAX_PARTICLES << endl;
fOutput << "MAX_SITES = " << MAX_SITES << endl;
fOutput << "P_S_FACTOR = " << P_S_FACTOR << endl;
fOutput << "L_SAMPLES = " << L_SAMPLES << endl;
fOutput.close();

exit(EXIT_SUCCESS);
}


long double hop_rate(int x)
{
```

```
  long double val;

  if(x!=0)
    val = (long double) ( 1.0 + (LAMBDA/(long double) x) );
  else
    val = 0.0;

  return val;
}

void partition(int startSites, int endSites)
{
  int i,j,k,m,n;
  long double fExpLogSum;

  cout << "startSites = " << startSites << "; endSites = "
       << endSites << endl;

if(startSites == 1){
  startSites++;
}

for(i=startSites;i<=endSites;i++){
  m = (i%2);  // use alternating rows to lower memory usage
  n = ((m+1)%2);
  lnPartitionArray[m][0] = 0.0;
  for(j=1;j<=MAX_PARTICLES+1;j++){
    fExpLogSum = 1.0;  // set to one before adding sum
    for(k=1;k<=j;k++){
      fExpLogSum += (marginalArray[k]
                     * exp( lnPartitionArray[n][j-k]
                            - lnPartitionArray[n][j] ));
    }
    lnPartitionArray[m][j] = lnPartitionArray[n][j]
                             + log(fExpLogSum);
  }
  cout << "Calculated Z for " << i << " sites." << endl;
```

```
  }

}
```

Appendix E

Code to Simulate the Non-conserving Zero-Range Process Model

```
//-----------------------------------------------------------------
// C++ code to simulate the homogeneous zero range
// process with creation and annihilation of particles
// and calculate the probability distribution
// of the number of particles at a site in the steady state.
//
//-----------------------------------------------------------------
// Written by Andrew Angel
// August 2004
// Last Update 1st Aug 2005
// Compiler: g++
//-----------------------------------------------------------------

#include <iostream.h>
#include <stdlib.h>
#include <stdio.h>
#include <fstream.h>
#include <math.h>
#include <string.h>
#include <strstream.h>
```

```
// global constants

const long double LAMBDA = 2.6;      //parameters for run
const unsigned int INITIAL_PARTICLE_NO = 1000;
const unsigned int MAX_PARTICLE_NO = 10000;
const unsigned int LATTICE_SIZE = 5000;
const unsigned int NO_MC_STEPS = 20000000;
const unsigned int OUTPUT_NUMBER = 1000;
const unsigned int AVERAGE_FACTOR = 400;
const unsigned int AVERAGE_NUMBER = OUTPUT_NUMBER*AVERAGE_FACTOR;
const int SEED = 0; // random num seeded from the system time if 0
const unsigned int RANDOM_FILL = 1;
// 0 -> start all particles on one site, random ICs otherwise
const unsigned int FILLED_SITE = 500;  // where filled site is.
const long double K_VAL = 3.0;
const long double S_VAL = 3.5;
const long double A_VAL = 1.0;
const long double DROLLNORM = 2.0+LAMBDA+pow(LATTICE_SIZE,-S_VAL);
const long double SNAP_NORM = 5.0/(long double)LATTICE_SIZE;

// global variables

unsigned int config[LATTICE_SIZE] = {0};   // holds system config
long double hopRateLookup[MAX_PARTICLE_NO + 1] = {0};
unsigned int particleNo = INITIAL_PARTICLE_NO;
unsigned int hopNum = 0;
unsigned int createNum = 0;
unsigned int annihilateNum = 0;
long double unnormalisedDistribution[MAX_PARTICLE_NO+1] = {0.0};
long double normalisedDistribution[MAX_PARTICLE_NO+1] = {0.0};
long double normalisation = 0.0;

// procedures and functions

void random_particle_fill(unsigned int, unsigned int);

void mc_update(unsigned int);
```

```
void create_look_up(unsigned int, unsigned int);

long double hop_rate(int x)
{
  long double val;

  if (x != 0){
    val = (long double) 1.0 + ( LAMBDA / (long double) x );
  }
  else{
    val = (long double) 0.0;
  }
  return val;
}

long double creation_rate(unsigned int, unsigned int, long double);
long double annihilation_rate(unsigned int, unsigned int,
                              long double, long double);

void add_to_unnormed_dist();

// begin main
main()
{
  // declare local variables

  unsigned int totalSteps = LATTICE_SIZE * NO_MC_STEPS;
  unsigned int updatesPerStep = LATTICE_SIZE;  // for readability
  unsigned int averageUpdates = updatesPerStep / AVERAGE_FACTOR;
  unsigned int i,j,k,l,m,n;   // counter variables for loops
  unsigned int totalParticles;
  char flNmStr[25];
  unsigned int outputInterval
                  = (int) ceil( (double) NO_MC_STEPS /
                                (double) OUTPUT_NUMBER );
```

```
unsigned int averageInterval
                = (int) ceil( (double) outputInterval /
                              (double) AVERAGE_FACTOR );

// open data streams
ofstream fPartNo;
ofstream fHopNo;
ofstream fCreateNo;
ofstream fAnnihilateNo;
ofstream fDist;
ofstream fData;
ofstream fSnapDist;

srand48(time(NULL));

// create the look up tables
create_look_up(LATTICE_SIZE, MAX_PARTICLE_NO);

// name and open files
fPartNo.open("particleno.dat");
fPartNo << 0 << " " << INITIAL_PARTICLE_NO << endl;
fHopNo.open("hopno.dat");
fHopNo << 0 << " " << hopNum << endl;
fCreateNo.open("createno.dat");
fCreateNo << 0 << " " << createNum << endl;
fAnnihilateNo.open("annihilateno.dat");
fAnnihilateNo << 0 << " " << annihilateNum << endl;

if (RANDOM_FILL == 0){  // random fill or all on one
  for (j=0 ; j < LATTICE_SIZE; j++){
    config[j] = 0;
  }
  config[FILLED_SITE] = INITIAL_PARTICLE_NO ;
}
else{
  random_particle_fill(INITIAL_PARTICLE_NO, LATTICE_SIZE);
}
```

```
// First loop to get to steady state
for(j=0;j<OUTPUT_NUMBER;j++){
  for(k=0;k<outputInterval;k++){
    for(l=0;l<updatesPerStep;l++){
      mc_update(LATTICE_SIZE);
    }
  }
  fPartNo << j+1 << " " << particleNo << endl;
  fHopNo << j+1 << " " << hopNum << endl;
  fCreateNo << j+1 << " " << createNum << endl;
  fAnnihilateNo << j+1 << " " << annihilateNum << endl;
}

// Second loop average the probability distribution
for(j=0;j<OUTPUT_NUMBER;j++){
  for(k=0;k<AVERAGE_FACTOR;k++){
    for(n=0;n<averageInterval;n++){
      for(l=0;l<updatesPerStep;l++){
        mc_update(LATTICE_SIZE);
      }
    }
    add_to_unnormed_dist();
  }
  fPartNo << j+1+OUTPUT_NUMBER << " " << particleNo << endl;
  fHopNo << j+1+OUTPUT_NUMBER << " " << hopNum << endl;
  fCreateNo << j+1+OUTPUT_NUMBER << " " << createNum << endl;
  fAnnihilateNo << j+1+OUTPUT_NUMBER << " " << annihilateNum
                << endl;
}

fDist.open("distribution.dat");
normalisation = 0.0;
for(m=0;m<=MAX_PARTICLE_NO;m++)
  normalisation += unnormalisedDistribution[m];
for(m=0;m<=MAX_PARTICLE_NO;m++){
  normalisedDistribution[m] = unnormalisedDistribution[m]
```

```
                                                    /(normalisation);
    fDist << m << " " << normalisedDistribution[m] << endl;
  }
  fDist.close();

  ostrstream name3(flNmStr, sizeof flNmStr);
  name3 << "final.data" << ends;
  fData.open(flNmStr);
  for(i=0;i<LATTICE_SIZE;i++){
    fData << i << " " << config[i] << endl;
  }
  fData.close();

  fPartNo.close();
  fHopNo.close();
  fCreateNo.close();
  fAnnihilateNo.close();

  // output stuff to a log file for error checking

  ofstream fLog;
  ostrstream name5(flNmStr, sizeof flNmStr);
  name5 << "output.log" << ends;
  fLog.open(flNmStr);

  fLog << "Lambda = " << LAMBDA << endl;
  fLog << "Lattice Size = " << LATTICE_SIZE << endl;
  fLog << "Initial Number of Particles = " << INITIAL_PARTICLE_NO
       << endl;
  fLog << "Maximum Number of Particles = " << MAX_PARTICLE_NO
       << endl;
  fLog << "Number of Monte Carlo Steps = " << NO_MC_STEPS << endl;
  fLog << "Number of Averages Taken = " << AVERAGE_NUMBER << endl;
  fLog << "SNAP_NORM = " << SNAP_NORM << endl;
  fLog << "Number of particle number points output = "
       << OUTPUT_NUMBER*2 << endl;
  fLog << "Output Interval = " << outputInterval << endl;
```

```
fLog << "Seed = " << SEED << endl;
fLog << "Random Fill status was: " << RANDOM_FILL << endl;
fLog << "Filled Site = " << FILLED_SITE << endl;
totalParticles = 0;
for(i=0;i<LATTICE_SIZE;i++){
  totalParticles = totalParticles + config[i];
}
fLog << "Totalled particles at end = " << totalParticles
     << endl;
fLog << "K_VAL = " << K_VAL << endl;
fLog << "A_VAL = " << A_VAL << endl;
fLog << "S_VAL = " << S_VAL << endl;
fLog << endl;
fLog << "Hop Rate: 1+LAMBDA/occupation" << endl;
fLog << "Creation Rate: 1/LATTICE_SIZE^S_VAL" << endl;
fLog << "Annihilation Rate : A_VAL*(occupation
                                    /totalParticleNo)^K_VAL"
     << endl;
  fLog << endl;
  fLog << "Total number of hops: " << hopNum << endl;
  fLog << "Total number of creations: " << createNum << endl;
  fLog << "Total number of annihilations: " << annihilateNum << endl;
  fLog.close();

  exit(EXIT_SUCCESS);
}


// procedures and functions in full if not already so above

void create_look_up(unsigned int latticeSize,
                    unsigned int maxParticles)
{
  int i = 0;

  for ( i = 0; i <= maxParticles ; i++){
    hopRateLookup[i] = hop_rate(i);
```

```
  }
}

void random_particle_fill(unsigned int particleNumber,
                          unsigned int latticeSize)
{
  unsigned int i;
  unsigned int randSite;

  for(i=0;i<latticeSize;i++){
    config[i] = 0;
  }

  for(i=0;i<particleNumber;i++){
    randSite = (unsigned int) floor( drand48()
                                     * (long double) latticeSize );
    config[randSite]++;
  }
  return;
}

void mc_update(unsigned int latticeSize)
{
  unsigned int site;
  long double diceRoll;
  long double diceRoll2;
  unsigned int destinationSite;
  long double c;

  site = (unsigned int) floor( drand48() * latticeSize );
  diceRoll = drand48()*DROLLNORM;

  c = creation_rate(LATTICE_SIZE,config[site],S_VAL);

  if(diceRoll < c){
    config[site] ++;
    particleNo ++;
```

```
    createNum ++;
  }
  else if(diceRoll > c
          && diceRoll < c +
          annihilation_rate(LATTICE_SIZE,config[site],K_VAL,A_VAL)){
    config[site] --;
    particleNo --;
    annihilateNum ++;
  }
  else if(diceRoll > c + 1.0 &&
          diceRoll < c + 1.0 + hopRateLookup[config[site]]){
    destinationSite = (unsigned int) floor(drand48()*latticeSize );
    config[site] = config[site] --;
    config[destinationSite] ++;
    hopNum ++;
  }
  return;
}
```

```
long double creation_rate(unsigned int noSites, unsigned int x,
                          long double sVal)
{
  long double val;

  val = 1/(pow(noSites,sVal));

  return val;
}

long double annihilation_rate(unsigned int noSites, unsigned int x,
                              long double kVal, long double aVal)
{
  long double val;

  val = aVal*pow((((long double) x)/((long double) noSites) ),kVal);
  val = val/aVal;
```

```
  return val;
}

void add_to_unnormed_dist()
{
  int i;

  for(i=0;i<LATTICE_SIZE;i++)
    unnormalisedDistribution[config[i]] += SNAP_NORM;

  return;
}
```

Appendix F

Code to Simulate the Non-Conserving Network Model Without Self Links

This code was written and supplied by Erel Levine.

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>

#define SWAP(a,b) {tmp=a;a=b;b=tmp;}

#define WIN 200.0
#define ARGPIC 8

#define MaxEdges 50000
#define MaxEdgesPerSite 5000

typedef struct __Edge {
  int v1, v2;
  int n1, n2;
  int id;
} Edge;

int tmp;
```

```
int *P;
double one;

int* degree;
Edge** edges;
Edge*** graph;

double b,c,theta;
int N, N0, M;

int checkedge(Edge* e, char* s) {
  if (e->v1 == e->v2)
    printf("%s: edge from vertex to itself.\n",s);
  if (graph[e->v1][e->n1]!=e) {
        printf("%s[1]: (%d,%d) %d [%d] %d [%d] %d\n",
               s,e->v1,e->v2,e->n1,degree[e->v1],
               e->n2,degree[e->v2],e->id);
        return(1);
  }
  if (graph[e->v2][e->n2]!=e) {
        printf("%s[2]: (%d,%d) %d [%d] %d [%d] %d\n",
               s,e->v1,e->v2,e->n1,degree[e->v1],
               e->n2,degree[e->v2],e->id);
        return(1);
  }
  if(edges[e->id]!=e) {
        printf("%s[E]: (%d,%d) %d [%d] %d [%d] %d\n",
               s,e->v1,e->v2,e->n1,degree[e->v1],
               e->n2,degree[e->v2],e->id);
    return(1);
  }
  return(0);
}

int addRandomBond(int i) {
  int j;
```

```
if (N>=MaxEdges) {
  fprintf(stderr,"Error: N > MaxEdges.\n");
  exit(1);
}
do {
  j = (int)(drand48()*M);
} while (j==i);
if (degree[i]>=MaxEdgesPerSite || degree[j]>=MaxEdgesPerSite) {
  fprintf(stderr,"Error: degree > MaxEdgesPerSite.\n");
  exit(1);
}
//  if (i>j)
//    SWAP(i,j);
edges[N] = malloc(sizeof(Edge));
edges[N]->v1 = i;
edges[N]->v2 = j;
graph[i][degree[i]] = edges[N];
  edges[N]->n1 = degree[i];
  graph[j][degree[j]] = edges[N];
  edges[N]->n2 = degree[j];
  edges[N]->id = N;
  N++; degree[i]++; degree[j]++;
}

int removeRandomBond(int i) {
  Edge* e;
  int j;

  if (degree[i]==0)
    return (-1);
  j = (int)(drand48()*degree[i]);
  e = graph[i][j];
  graph[e->v1][e->n1] = graph[e->v1][degree[e->v1]-1];
  if (e->v1==graph[e->v1][e->n1]->v1)
    graph[e->v1][e->n1]->n1 = e->n1;
  else
```

```
    graph[e->v1][e->n1]->n2 = e->n1;
  degree[e->v1]--;
  graph[e->v2][e->n2] = graph[e->v2][degree[e->v2]-1];
  if (e->v2 == graph[e->v2][e->n2]->v2)
    graph[e->v2][e->n2]->n2 = e->n2;
  else
    graph[e->v2][e->n2]->n1 = e->n2;
  degree[e->v2]--;
  edges[e->id] = edges[N-1];
  edges[e->id]->id = e->id;
  N--;
  free(e);
  return(0);
}

int moveRandomBond(int i) {
  Edge* e;
  int j,k;

  k = (int)(drand48()*degree[i]);

  e = graph[i][k];
  do {
    j = (int)(drand48()*M);
  } while (j==e->v1 || j==e->v2);
  if (e->v1 == i && e->n1==k) {
    graph[e->v1][e->n1] = graph[e->v1][degree[e->v1]-1];
    if (e->v1==graph[e->v1][e->n1]->v1)
      graph[e->v1][e->n1]->n1 = e->n1;
    else
      graph[e->v1][e->n1]->n2 = e->n1;
    e->v1 = j;
    e->n1 = degree[j];
  } else if (e->v2 == i && e->n2==k) {
    graph[e->v2][e->n2] = graph[e->v2][degree[e->v2]-1];
    if (e->v2 == graph[e->v2][e->n2]->v2)
      graph[e->v2][e->n2]->n2 = e->n2;
```

```
    else
      graph[e->v2][e->n2]->n1 = e->n2;
    e->v2 = j;
    e->n2 = degree[j];
  }
  else{
    printf("Error in link update\n");
    printf("i=%i, j=%i, k=%i, e->v1=%i, e->n1=%i,
            e->v2=%i, e->n2=%i\n",
           i,j,k,e->v1,e->n1,e->v2,e->n2);
    exit(1);
  }
  graph[j][degree[j]]=e;
  degree[j]++;
  degree[i]--;
}

void simulate () {
int s1, s2, t, i, j;
Edge* e;
double r, w;

for (s1=0;s1<WIN;s1++) {
  for (s2=0;s2<WIN;s2++) {
    for (t=0;t<M;t++) {
      i = (int)(drand48()*M);
      r = drand48()*one;
      w = pow((double)(degree[i])/M,theta);
      if (r<c) {
        addRandomBond(i);
      } else if (r>c && r<(w+c)) {
        removeRandomBond(i);
      } else if (degree[i]>0 && r>(c+1.0) &&
                 r<(c+2.0 +b/(double)(degree[i]))) {
        moveRandomBond(i);
      }
    }
```

```
    }
    for (t=0;t<M;t++)
      if (degree[t]<N0)
        P[degree[t]]++;
  }
}

int main(int argc, char* argv[]) {
  int swp,i,j,k;
  char fname[255];
  FILE *pfile,*configfile, *nfile;
  double s;

  if (argc<ARGPIC) {
    printf("Usage: %s M Phi b s k swp prefix [picflag]\n",argv[0]);
    exit(1);
  }
  srand48(time(NULL));

  /* variables */
  M      = atoi(argv[1]);
  N0     = (int)(atof(argv[2])*M);
  b      = atof(argv[3]);
  s      = atof(argv[4]);
  c      = pow(M,-s);
  theta  = atof(argv[5]);
  one    = (1.0+b) + c + 1.0;
  swp    = atoi(argv[6]);

  strcpy(fname,argv[7]);
  pfile = fopen(fname,"w");
  strcpy(fname,argv[7]);
  strcat(fname,".N");
  nfile = fopen(fname,"w");
  if (argc>ARGPIC) {
    strcpy(fname,argv[7]);
    strcat(fname,".pic");
```

```
  configfile = fopen(fname,"w");
}

/* Memory allocation */
degree =   (int*)   calloc(M,sizeof(int));
edges  =   (Edge**) calloc(MaxEdges,sizeof(Edge*));
graph  =   (Edge***)calloc(M,sizeof(Edge**));
for (i=0;i<M;i++)
  graph[i]=(Edge**) calloc(MaxEdgesPerSite,sizeof(Edge*));

/* Random initial configuration */
for (N=0;N<N0;) {
  addRandomBond((int)(drand48()*M));
}
N0 *= 3;
P = (int*)calloc(N0,sizeof(int));

fprintf(stderr,"M=%d N0=%d b=%lf c=%lf k=%lf\n",M,N,b,c,theta);
```

```
for (i=0;i<swp;i++) {
  simulate();
  if (argc>ARGPIC) {
    for (k=0;k<M;k++)
      fprintf(configfile,"%d ",(int)degree[k]);
    fprintf(configfile,"\n");
  }
  fprintf(nfile,"%d\n",N);
  fflush(nfile);
  for (k=0;k<N0;k++) {
    if (P[k]>0) {
      fprintf(pfile,"%d %f\t",k,P[k]/WIN);
      P[k]=0;
    }
  }
  fprintf(pfile,"\n");
}
```

```
  if (argc>ARGPIC)
    fclose(configfile);
  fclose(pfile);
}
```

Appendix G

Code to Simulate the Non-Conserving Network Model with Self Links Allowed

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>

#define SWAP(a,b) {tmp=a;a=b;b=tmp;}

#define WIN 200.0
#define ARGPIC 8

#define MaxEdges 50000
#define MaxEdgesPerSite 5000

typedef struct __Edge {
  int v1, v2;
  int n1, n2;
  int id;
} Edge;

int tmp;

int *P;
```

```
double one;

int* degree;
Edge** edges;
Edge*** graph;

double b,c,theta;
int N, N0, M;

int createNo = 0;
int annihilateNo = 0;

int checkedge(Edge* e, char* s) {
  if (e->v1 == e->v2)
    printf("%s: edge from vertex to itself.\n",s);
  if (graph[e->v1][e->n1]!=e) {
        printf("%s[1]: (%d,%d) %d [%d] %d [%d] %d\n",
               s,e->v1,e->v2,e->n1,degree[e->v1],
               e->n2,degree[e->v2],e->id);
        return(1);
   }
   if (graph[e->v2][e->n2]!=e) {
        printf("%s[2]: (%d,%d) %d [%d] %d [%d] %d\n",
               s,e->v1,e->v2,e->n1,degree[e->v1],
               e->n2,degree[e->v2],e->id);
        return(1);
   }
   if(edges[e->id]!=e) {
        printf("%s[E]: (%d,%d) %d [%d] %d [%d] %d\n",
               s,e->v1,e->v2,e->n1,degree[e->v1],
               e->n2,degree[e->v2],e->id);
     return(1);
   }
   return(0);
}

int addRandomBond(int i) {
```

```
int j;

if (N>=MaxEdges) {
  fprintf(stderr,"Error: N > MaxEdges.\n");
  exit(1);
}
  j = (int)(drand48()*M);
if ( degree[i]>=MaxEdgesPerSite || degree[j]>=MaxEdgesPerSite
     || (i==j && degree[i]>=(MaxEdgesPerSite-1)) ) {
  fprintf(stderr,"Error: degree > MaxEdgesPerSite.\n");
  exit(1);
}
//  if (i>j)
//    SWAP(i,j);
edges[N] = malloc(sizeof(Edge));
edges[N]->v1 = i;
edges[N]->v2 = j;
graph[i][degree[i]] = edges[N];
  edges[N]->n1 = degree[i];
  degree[i]++;
  graph[j][degree[j]] = edges[N];
  edges[N]->n2 = degree[j];
  edges[N]->id = N;
  N++; degree[j]++;
  // for testing
  // printf("Added a bond.\n");
  createNo+=2;
}

int removeRandomBond(int i) {
  Edge* e;
  int j;

  // for testing
  // printf("Started trying to remove a bond.\n");

  if (degree[i]==0)
```

```
    return (-1);
  j = (int)(drand48()*degree[i]);
  e = graph[i][j];
  graph[e->v1][e->n1] = graph[e->v1][degree[e->v1]-1];
  if ( (e->v1==graph[e->v1][e->n1]->v1) &&
       (graph[e->v1][e->n1]->n1==(degree[e->v1]-1)) )
    graph[e->v1][e->n1]->n1 = e->n1;
  else
    graph[e->v1][e->n1]->n2 = e->n1;
  degree[e->v1]--;
  graph[e->v2][e->n2] = graph[e->v2][degree[e->v2]-1];
  if ( (e->v2==graph[e->v2][e->n2]->v2) &&
       (graph[e->v2][e->n2]->n2==(degree[e->v2]-1)) )
    graph[e->v2][e->n2]->n2 = e->n2;
  else
    graph[e->v2][e->n2]->n1 = e->n2;
  degree[e->v2]--;
  edges[e->id] = edges[N-1];
  edges[e->id]->id = e->id;
    N--;
    free(e);
    //for testing
    //printf("Removed a bond.\n");
    annihilateNo+=2;
    return(0);
}

int moveRandomBond(int i) {
    Edge* e;
    int j,k;

    // for testing
    // printf("started trying to move a bond.\n");

    k = (int)(drand48()*degree[i]);
    e = graph[i][k];
    j = (int)(drand48()*M);
```

```
if (degree[j]>=MaxEdgesPerSite){
  fprintf(stderr,"Error: degree > MaxEdgesPerSite.\n ");
  exit(1);
}

if ( (e->v1==i) && (e->n1==k) ) {
  // for testing
  // printf("e->v1==i&&e->n1==k\n");
  graph[e->v1][e->n1] = graph[e->v1][degree[e->v1]-1];
  if ( (e->v1==graph[e->v1][e->n1]->v1) &&
       (graph[e->v1][e->n1]->n1==(degree[e->v1]-1)) )
    graph[e->v1][e->n1]->n1 = e->n1;
  else
    graph[e->v1][e->n1]->n2 = e->n1;
  degree[i]--;
  e->v1 = j;
  e->n1 = degree[j];
} else if ( (e->v2==i) && (e->n2==k) ){
  // for testing
  // printf("e->v2==i&&e->n2==k\n");
  graph[e->v2][e->n2] = graph[e->v2][degree[e->v2]-1];
  if ( (e->v2==graph[e->v2][e->n2]->v2) &&
       (graph[e->v2][e->n2]->n2==(degree[e->v2]-1)) )
    graph[e->v2][e->n2]->n2 = e->n2;
  else
    graph[e->v2][e->n2]->n1 = e->n2;
  degree[i]--;
  e->v2 = j;
  e->n2 = degree[j];
} else {
  printf("Error in link update\n");
  printf("i=%i, j=%i, k=%i, e->v1=%i, e->n1=%i,
          e->v2=%i, e->n2=%i\n"
         ,i,j,k,e->v1,e->n1,e->v2,e->n2);
  exit(1);
}
```

```
  graph[j][degree[j]]=e;
  degree[j]++;
  // for testing
  // printf("Moved a bond.\n");
}

void simulate () {
  int s1, s2, t, i, j;
  Edge* e;
  double r, w;

  for (s1=0;s1<WIN;s1++) {
    for (s2=0;s2<WIN;s2++) {
      for (t=0;t<M;t++) {
        i = (int)(drand48()*M);
        r = drand48()*one;
        w = pow((double)(degree[i])/M,theta);
        if (r<c) {
          addRandomBond(i);
        } else if (r>c && r<(w+c)) {
          removeRandomBond(i);
        } else if (degree[i]>0 && r>(c+1.0) &&
                  r<(c+2.0 +b/(double)(degree[i]))) {
          moveRandomBond(i);
        }
      }
    }
    for (t=0;t<M;t++)
      if (degree[t]<N0)
        P[degree[t]]++;
  }
}

int main(int argc, char* argv[]) {
  int swp,i,j,k;
  char fname[255];
```

```
FILE *pfile,*configfile, *nfile, *afile, *cfile;
double s;

if (argc<ARGPIC) {
  printf("Usage: %s M Phi b s k swp prefix [picflag]\n",argv[0]);
  exit(1);
}
srand48(time(NULL));

/* variables */
M      = atoi(argv[1]);
N0     = (int)(atof(argv[2])*M);
b      = atof(argv[3]);
s      = atof(argv[4]);
c      = pow(M,-s);
theta  = atof(argv[5]);
one    = (1.0+b) + c + 1.0;
swp    = atoi(argv[6]);

strcpy(fname,argv[7]);
pfile = fopen(fname,"w");
strcpy(fname,argv[7]);
strcat(fname,".N");
nfile = fopen(fname,"w");
if (argc>ARGPIC) {
  strcpy(fname,argv[7]);
  strcat(fname,".pic");
  configfile = fopen(fname,"w");
}
afile = fopen("annihilateno.dat","w");
cfile = fopen("createno.dat","w");

/* Memory allocation */
degree =   (int*)   calloc(M,sizeof(int));
edges  =   (Edge**) calloc(MaxEdges,sizeof(Edge*));
graph  =   (Edge***)calloc(M,sizeof(Edge**));
for (i=0;i<M;i++)
```

```
  graph[i]=(Edge**) calloc(MaxEdgesPerSite,sizeof(Edge*));

/* Random initial configuration */
for (N=0;N<N0;) {
  addRandomBond((int)(drand48()*M));
}
N0 *= 3;
P = (int*)calloc(N0,sizeof(int));

fprintf(stderr,"M=%d N0=%d b=%lf c=%lf k=%lf\n",M,N,b,c,theta);


for (i=0;i<swp;i++) {
  simulate();
  if (argc>ARGPIC) {
    for (k=0;k<M;k++)
      fprintf(configfile,"%d ",(int)degree[k]);
    fprintf(configfile,"\n");
  }
    fprintf(nfile,"%d\n",N);
    fflush(nfile);
    for (k=0;k<N0;k++) {
      if (P[k]>0) {
        fprintf(pfile,"%d %f\t",k,P[k]/WIN);
        P[k]=0;
      }
    }
    fprintf(afile,"%i\n",annihilateNo);
    fflush(afile);
    fprintf(cfile,"%i\n",createNo);
    fprintf(pfile,"\n");
  }

  if (argc>ARGPIC)
    fclose(configfile);
  fclose(pfile);
}
```

Appendix H

Code to Simulate the Multi-Species Zero-Range Process Directed, Weighted Network Model

```
//-----------------------------------------------------------------------
// C++ code to simulate the homogeneous multi-species
// zero-range process model of a directed, weighted network
// This version calculates site and column steady-state
// probability distributions
// This is the u_c(l) = u_0  if l<=L   version
//                      1+lambda/l  if l>L
// u_0 <= 1 + b_c/L is required to see
// a condensation transition
// Also outputs the in and out degree distributions for
// network comparison
//-----------------------------------------------------------------------
// Written by Andrew Angel
// Mar 2005 (original Aug 2004)
// Last Update 16th Jun 2005
// Compiler: g++
//-----------------------------------------------------------------------
```

```
#include <iostream.h>
#include <stdlib.h>
#include <stdio.h>
#include <fstream.h>
#include <math.h>
#include <string.h>
#include <strstream.h>
// #include "rng.h"

// global constants

const long double B_S = 0.0;
const long double B_C = 1.05;
const unsigned int PARTICLES_PER_ROW = 2000;
const unsigned int NO_COLUMNS = 100;
const unsigned int NO_ROWS = 100;
const long double U_0 = 1.5;
const unsigned int TOTAL_SITES = (NO_ROWS * NO_COLUMNS);
const unsigned int TOTAL_PARTICLES = (PARTICLES_PER_ROW * NO_ROWS);
const unsigned int NO_MC_STEPS = 20000000;  //
const unsigned int OUTPUT_NUMBER = 5000;
const unsigned int AVERAGE_FACTOR = 100;
const unsigned int AVERAGE_NUMBER = OUTPUT_NUMBER*AVERAGE_FACTOR;
// how many samples will be taken
const int SEED = 0; // random num seeded from system time if 0
const unsigned int RANDOM_FILL = 0;
// 0 -> start all particles on one site; 1->half on one site
// 2-> RFN*L on one site
const long double RFN = 0.9;
//fraction of particles to go on the slow site if RANDOM_FILL == 2
const unsigned int FILLED_SITE = 50;  // where the filled site is.
const long double SNAP_NORM = (long double) NO_ROWS;

// global variables

unsigned int config[NO_ROWS][NO_COLUMNS] = {0};
// holds the system's configuration
```

```
unsigned int columnConfig[NO_COLUMNS] = {0};
long double siteHopRateLookup[PARTICLES_PER_ROW + 1] = {0};
long double columnHopRateLookup[TOTAL_PARTICLES + 1] = {0};
unsigned int hopNum = 0;

long double unnormedDistribution[NO_ROWS][PARTICLES_PER_ROW+1]
                                                    = {0.0};
long double normedDistribution[NO_ROWS][PARTICLES_PER_ROW+1] = {0.0};
long double unnormedSiteDistribution[PARTICLES_PER_ROW+1] = {0.0};
long double normedSiteDistribution[PARTICLES_PER_ROW+1] = {0.0};
long double unnormedColumnDistribution[TOTAL_PARTICLES+1] = {0.0};
long double normedColumnDistribution[TOTAL_PARTICLES+1] = {0.0};
long double unnormedInColumnDistribution[NO_COLUMNS]
                                        [PARTICLES_PER_ROW+1] = {0.0};
long double normedInColumnDistribution[NO_COLUMNS]
                                      [PARTICLES_PER_ROW+1] = {0.0};
long double unnormedInDegreeDistribution[NO_ROWS+1] = {0.0};
long double normedInDegreeDistribution[NO_ROWS+1] = {0.0};
long double unnormedOutDegreeDistribution[NO_COLUMNS+1] = {0.0};
long double normedOutDegreeDistribution[NO_COLUMNS+1] = {0.0};


// procedures and functions

void random_particle_fill(unsigned int, unsigned int, unsigned int);

void mc_update(unsigned int);

void create_site_look_up(unsigned int);
void create_column_look_up(unsigned int);

long double site_hop_rate(int x)
{
  long double val;

  if (x != 0){
    val = (long double) 1.0 + ( B_S / (long double) x );
```

```
  }
  else{
    val = (long double) 0.0;
  }
  if( U_0 > (1+B_C/(long double)NO_COLUMNS) )
    val = val/( (1.0 + B_S)*(U_0) );
  else
    val = val/( (1.0+B_S)*(1+B_C/(long double)NO_COLUMNS) );
  return val;
}

long double column_hop_rate(int x)
{
  long double val;

  if (x != 0){
    if ( x <= NO_COLUMNS ){
      val = U_0;
    }
    else{
      val = 1.0 + ( B_C / (long double) x );
    }
  }
  else{
    val = 0.0;
  }
  return val;
}

void add_to_unnormed_dists();

// begin main
main()
{
  // declare local variables

  unsigned int totalSteps = TOTAL_SITES * NO_MC_STEPS;
```

```
unsigned int totalParticles = 0;
unsigned int updatesPerStep = TOTAL_SITES;  // for readability
unsigned int averageUpdates = updatesPerStep / AVERAGE_FACTOR;
unsigned int i,j,k,l,m,n;   // counter variables for loops

char flNmStr[25];

unsigned int outputInterval =
                  (int) ceil( (double) NO_MC_STEPS /
                              (double) OUTPUT_NUMBER );
unsigned int averageInterval =
                  (int) ceil( (double) outputInterval /
                              (double) AVERAGE_FACTOR );

long double norm;

// open data streams
ofstream fDist;
ofstream fSiteDist;
ofstream fColDist;
ofstream fInColDist;
ofstream fInitial;
ofstream fFinal;
ofstream fInDegree;
ofstream fOutDegree;
ofstream fSeedColOcc;

srand48(time(NULL)); // seed the random number generator

// create the look up tables
create_site_look_up(PARTICLES_PER_ROW);
create_column_look_up(TOTAL_PARTICLES);

if (RANDOM_FILL == 0){        //random fill or all on one
  for (i=0; i < NO_ROWS; i++){
    for (j=0 ; j < NO_COLUMNS; j++){
      config[i][j] = 0;
```

```
      }
      config[i][FILLED_SITE] = PARTICLES_PER_ROW;
      columnConfig[FILLED_SITE] += PARTICLES_PER_ROW;
    }
  }
  else if (RANDOM_FILL == 1){
    for(j=0;j<NO_ROWS;j++)
      random_particle_fill(PARTICLES_PER_ROW/2,NO_COLUMNS,j);
    for(i=0;i<NO_ROWS;i++){
      config[i][FILLED_SITE] += (PARTICLES_PER_ROW
                                 - PARTICLES_PER_ROW/2);
      columnConfig[FILLED_SITE] += (PARTICLES_PER_ROW
                                    - PARTICLES_PER_ROW/2);
    }
  }
  else if (RANDOM_FILL == 2){
    for(j=0;j<NO_ROWS;j++)
      random_particle_fill(
              (int)ceil((1.0-RFN)*PARTICLES_PER_ROW),NO_COLUMNS,j);
    for(i=0;i<NO_ROWS;i++){
      config[i][FILLED_SITE] +=
        (PARTICLES_PER_ROW - (int)ceil((1.0-RFN)*PARTICLES_PER_ROW));
      columnConfig[FILLED_SITE] +=
        (PARTICLES_PER_ROW - (int)ceil((1.0-RFN)*PARTICLES_PER_ROW));
    }
  }
  else{
    for (j=0; j < NO_ROWS; j++){
      random_particle_fill(PARTICLES_PER_ROW, NO_COLUMNS, j);
    }
  }

  // output the initial configuration
  fInitial.open("initial.data");
  for(i=0;i<NO_ROWS;i++){
    for(j=0;j<NO_COLUMNS;j++){
      fInitial << i << " " << j << " " << config[i][j] << endl;
```

```
  }
}
fInitial.close();

// open the output file for the size of the condensate
fSeedCol0cc.open("seedcolno.dat");

// first loop to get to steady state
// on the seeded site
for(j=0;j<OUTPUT_NUMBER;j++){
  for(k=0;k<AVERAGE_FACTOR;k++){
    for(n=0;n<averageInterval;n++){
      for(l=0;l<updatesPerStep;l++){
        mc_update(TOTAL_SITES);
      }
    }
    fSeedCol0cc << ( j*AVERAGE_FACTOR + k)  << " "
                << columnConfig[FILLED_SITE] << endl;
  }
}

// second loop take samples and get the probability distribution
for(j=0;j<OUTPUT_NUMBER;j++){
  for(k=0;k<AVERAGE_FACTOR;k++){
    for(n=0;n<averageInterval;n++){
      for(l=0;l<updatesPerStep;l++){
        mc_update(TOTAL_SITES);
      }
    }
    add_to_unnormed_dists();
    fSeedCol0cc << ( j*AVERAGE_FACTOR + k +
                     AVERAGE_FACTOR-1
                     + (OUTPUT_NUMBER-1)*(AVERAGE_FACTOR) )
                << " " << columnConfig[FILLED_SITE] << endl;
  }
}
fSeedCol0cc.close();
```

```
// output the final configuration
fFinal.open("final.data");
for(i=0;i<NO_ROWS;i++){
  for(j=0;j<NO_COLUMNS;j++){
    fFinal << i << " " << j << " " << config[i][j] << endl;
  }
}
fFinal.close();

// normalise the unnormalised distributions
for(i=0;i<NO_ROWS;i++){
  norm = 0;
  for(j=0;j<=PARTICLES_PER_ROW;j++){
    norm += unnormedDistribution[i][j];
  }
  for(j=0;j<=PARTICLES_PER_ROW;j++){
    normedDistribution[i][j] =
                ( (long double) unnormedDistribution[i][j]
                                / (long double) norm );
  }
}
norm = 0;
for(j=0;j<=PARTICLES_PER_ROW;j++){
  norm += unnormedSiteDistribution[j];
}
for(j=0;j<=PARTICLES_PER_ROW;j++){
  normedSiteDistribution[j] =
              ( (long double) unnormedSiteDistribution[j]
                              / (long double) norm );
}
norm = 0;
for(j=0;j<=TOTAL_PARTICLES;j++){
  norm += unnormedColumnDistribution[j];
}
for(j=0;j<=TOTAL_PARTICLES;j++){
  normedColumnDistribution[j] =
```

```
                ( (long double) unnormedColumnDistribution[j]
                                        / (long double) norm );
}
for(i=0;i<NO_COLUMNS;i++){
  norm = 0;
  for(j=0;j<=PARTICLES_PER_ROW;j++){
    norm += unnormedInColumnDistribution[i][j];
  }
  for(j=0;j<=PARTICLES_PER_ROW;j++){
    normedInColumnDistribution[i][j] =
      ( (long double) unnormedInColumnDistribution[i][j] /
                                          (long double) norm );
  }
}
norm = 0;
for(i=0;i<=NO_COLUMNS;i++){
  norm += unnormedOutDegreeDistribution[i];
}
for(i=0;i<=NO_COLUMNS;i++){
  normedOutDegreeDistribution[i] =
      ( (long double) unnormedOutDegreeDistribution[i] /
                                        (long double) norm );
}
norm = 0;
for(i=0;i<=NO_ROWS;i++){
  norm += unnormedInDegreeDistribution[i];
}
for(i=0;i<=NO_ROWS;i++){
  normedInDegreeDistribution[i] =
      ( (long double) unnormedInDegreeDistribution[i] /
                                      (long double) norm );
}

fDist.open("rowdistributions3d.dat");
for(i=0;i<NO_ROWS;i++){
  for(j=0;j<=PARTICLES_PER_ROW;j++){
    fDist << i << " " << j << " " << normedDistribution[i][j]
```

```
            << endl;
    }
  }
  fDist.close();
  fSiteDist.open("sitedistribution.dat");
  for(j=0;j<=PARTICLES_PER_ROW;j++){
   fSiteDist << j << " " << normedSiteDistribution[j] << endl;
  }
  fSiteDist.close();
  fColDist.open("columndistribution.dat");
  for(j=0;j<=TOTAL_PARTICLES;j++){
    fColDist << j << " " << normedColumnDistribution[j] << endl;
  }
  fColDist.close();
  fInColDist.open("incolumndistributions3d.dat");
  for(i=0;i<NO_COLUMNS;i++){
    for(j=0;j<=PARTICLES_PER_ROW;j++){
      fInColDist << i << " " << j << " "
        << normedInColumnDistribution[i][j] << endl;
    }
  }
  fInColDist.close();
  fOutDegree.open("outdegreedistribution.dat");
  for(i=0;i<=NO_COLUMNS;i++)
    fOutDegree << i << " " << normedOutDegreeDistribution[i] << endl;
  fOutDegree.close();
  fInDegree.open("indegreedistribution.dat");
  for(i=0;i<=NO_ROWS;i++)
    fInDegree << i << " " << normedInDegreeDistribution[i] << endl;
  fInDegree.close();

  // Output stuff to a log file for error checking

  ofstream fLog;

  fLog.open("output.log");
```

```
fLog << "B_S " << B_S << endl;
fLog << "B_C " << B_C << endl;
fLog << "U_0 " << U_0 << endl;
fLog << "NO_ROWS = " << NO_ROWS << endl;
fLog << "NO_COLUMNS = " << NO_COLUMNS << endl;
fLog << "PARTICLES_PER_ROW = " << PARTICLES_PER_ROW << endl;
fLog << "TOTAL_PARTICLES = " << TOTAL_PARTICLES << endl;
fLog << "Number of Monte Carlo Steps = " << NO_MC_STEPS << endl;
fLog << "Number of Outputs = " << OUTPUT_NUMBER << endl;
fLog << "Output Interval = " << outputInterval << endl;
fLog << "AVERAGE_FACTOR = " << AVERAGE_FACTOR << endl;
fLog << "AVERAGE_NUMBER = " << AVERAGE_NUMBER << endl;
fLog << "averageInterval = " << averageInterval << endl;
fLog << "Seed = " << SEED << endl;
fLog << "Random Fill status was: " << RANDOM_FILL << endl;
fLog << "Filled Site = " << FILLED_SITE << endl;
fLog << "Fraction started on filled site, RFN = " << RFN << endl;
totalParticles = 0;
for(i=0;i<NO_COLUMNS;i++){
  for(j=0;j<NO_ROWS;j++){
    totalParticles += config[i][j];
  }
}
fLog << "Totalled particles at end = " << totalParticles << endl;
fLog << "Site Hop Rate: 1+B_S/occupation" << endl;
fLog << endl;
fLog << "Column Hop Rate: U_0 l<=L; 1+B_C/l l>L" << endl;
fLog << endl;
fLog << "Total number of hops: " << hopNum << endl;
fLog.close();

// output the values of the site and column hop-rates
ofstream fHopSiteOut;
ofstream fColSiteOut;

fHopSiteOut.open("hopsite.dat");
for(i=0;i<=PARTICLES_PER_ROW;i++){
```

```
    fHopSiteOut << i << " " << siteHopRateLookup[i] << endl;
  }
  fHopSiteOut.close();
  fColSiteOut.open("hopcol.dat");
  for(i=0;i<=TOTAL_PARTICLES;i++){
    fColSiteOut << i << " " << columnHopRateLookup[i] << endl;
  }
  fColSiteOut.close();

  exit(EXIT_SUCCESS);

}


// procedures and functions in full if not already so above

void create_site_look_up(unsigned int maxParticles)
{
  int i = 0;Dorogovtsev, Mendes Samukhin

  for ( i = 0; i <= maxParticles ; i++){
    siteHopRateLookup[i] = site_hop_rate(i);
  }
}

void create_column_look_up(unsigned int maxParticles)
{
  int i = 0;

  for ( i = 0; i <= maxParticles ; i++){
    columnHopRateLookup[i] = column_hop_rate(i);
  }
}

void random_particle_fill(unsigned int particleNumber,
                          unsigned int rowSize, unsigned int rowIndex)
{
```

```
  unsigned int i;
  unsigned int randSite;

  for(i=0;i<rowSize;i++){
    config[rowIndex][i] = 0;
  }

  for(i=0;i<particleNumber;i++){
    randSite = (unsigned int) floor(drand48()
                                    * (long double) rowSize);
    config[rowIndex][randSite]++;
    columnConfig[randSite]++;
  }
  return;
}

void mc_update(unsigned int latticeSize)
{
unsigned int site[2];
unsigned int siteRoll;
long double diceRoll;
long double diceRoll2;
unsigned int destinationSite;

siteRoll = (unsigned int) floor(drand48()*TOTAL_SITES);
site[0] = siteRoll/NO_COLUMNS;  // want integer division here
site[1] = siteRoll%NO_COLUMNS;

diceRoll = drand48();

if( diceRoll < (siteHopRateLookup[config[site[0]][site[1]]]
                * columnHopRateLookup[columnConfig[site[1]]]) ) {
  destinationSite = (unsigned int) floor(drand48()*NO_COLUMNS);
  config[site[0]][site[1]] --;
  config[site[0]][destinationSite] ++;
  columnConfig[site[1]] --;
  columnConfig[destinationSite] ++;
```

```
    hopNum ++;
  }
  return;
}


void add_to_unnormed_dists()
{
  int i,j,k;
  int degree;

  for(i=0;i<NO_ROWS;i++){
    degree = 0;
    for(j=0;j<NO_COLUMNS;j++){
      unnormedDistribution[i][config[i][j]] += 1.0/SNAP_NORM;
      unnormedSiteDistribution[config[i][j]] += 1.0/SNAP_NORM;
      if(config[i][j] != 0)
        degree ++;
    }
    unnormedOutDegreeDistribution[degree] += 1.0/SNAP_NORM;
  }

  for(j=0;j<NO_COLUMNS;j++){
    unnormedColumnDistribution[columnConfig[j]] += 1.0/SNAP_NORM;
  }

  for(i=0;i<NO_COLUMNS;i++){
    degree = 0;
    for(j=0;j<NO_ROWS;j++){
      unnormedInColumnDistribution[i][config[j][i]] += 1.0/SNAP_NORM;
      if(config[j][i] != 0)
        degree ++;
    }
    unnormedInDegreeDistribution[degree] += 1.0/SNAP_NORM;
  }

}
```

Bibliography

[1] G.L. Vasconcelos, *A guided walk down wall street: an introduction to econophysics*, Braz. J. Phys. **34**, 1039 (2004)

[2] F. Spitzer, *Interaction of Markov processes*, Adv. Math. **5**, 246 (1970)

[3] M.R. Evans, *Phase transitions in one-dimensional nonequilibrium systems*, Braz. J. Phys. **30**, 42 (2000)

[4] M.R. Evans and T. Hanney, *Nonequilibrium statistical mechanics of the zero-range process and related models*, J. Phys. A **38**, R195 (2005)

[5] D.W. Stroock, *An Introduction to Markov Processes (Graduate Texts in Mathematics 230)*, (Springer-Verlag, Berlin Heidelberg, 2005)

[6] R. Albert and A.-L. Barabási, *Statistical mechanics of complex networks*, Rev. Mod. Phys. **74**, 47 (2002)

[7] S.N. Dorogovtsev and J.F.F. Mendes, *Evolution of networks*, Adv. in Phys. **51**, 1079 (2002)

[8] S.N. Dorogovtsev and J.F.F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW*, (Oxford University Press, Oxford) (2003)

[9] M.E.J. Newman, *The structure and function of complex networks*, SIAM Review **45**, 167-256, (2003)

[10] H.J. Weber and G.B. Arfken, *Essential Mathematical Methods for Physicists*, (Elsevier Academic Press, 2004)

[11] F.E. Beichelt and L.P. Fatti, *Stochastic Processes and their Applications*, (Taylor & Francis, London, 2002)

[12] W. Feller, *An Introduction to Probability Theory and its Applications, Vol. I*, Third Edition, (John Wiley & sons, New York, 1968)

[13] E.D. Andjel, *Invariant measures for the zero-range process*, Ann. Probab. **10**, 525 (1982)

[14] K. Huang, *Statistical Mechanics*, Second Edition, (John Wiley & Sons, Inc., 1987)

[15] O.J. O'Loan, M.R. Evans and M.E. Cates, *Jamming transition in a homogeneous, one-dimensional system: The bus route model*, Phys. Rev. E **58**, 1404 (1998)

[16] C. Godrèche, *Dynamics of condensation in zero-range processes*, J. Phys. A **36**, 6313 (2003)

[17] S. Großkinsky, G.M. Schütz and H. Spohn, *Condensation in the zero-range process: stationary and dynamical properties*, J. Stat. Phys. **113**, 389 (2003)

[18] N. Bleistein and R.A. Handelsman, *Asymptotic Expansions of Integrals*, (Holt, Rinehart and Winston, New York, 1975)

[19] M.R. Evans and T. Hanney, *Phase transition in two-species zero-range process*, J. Phys. A **36**, L441 (2003)

[20] T. Hanney and M.R. Evans, *Condensation transition in a two-species zero-range process*, Phys. Rev. E **69**, 016107 (2004)

[21] G.M. Schütz, *Critical phenomena and universal dynamics in one-dimensional driven diffusive systems with two species of particle*, J. Phys. A **36**, R339 (2003)

[22] S. Großkinsky and H. Spohn, *Stationary measures and hydrodynamics of zero range processes with several species of particles*, Bull. Braz. Math. Soc. **34**, 489 (2003)

[23] S. Großkinsky and T. Hanney, *Coarsening dynamics in a two-species zero-range process*, cond-mat/0412593

[24] M.R. Evans, *Exact steady states of disordered hopping particle models with parallel and ordered sequential dynamics*, J. Phys. A: Math. Gen. **30**, 5669 (1997)

[25] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C, The Art of Scientific Computing*, Second Edition, (Cambridge University Press, 1992)

[26] J.M.J. van Leeuwen and A. Kooiman, *The drift velocity in the Rubinstein-Duke model for electrophoresis*, Physica A **184**, 79 (1992)

[27] J.M. Carlson, E.R. Grannan and G.H. Swindle, *Self-organizing systems at finite driving rates*, Phys. Rev. E **47**, 93 (1993)

[28] F. Ritort, *Glassiness in a model without energy barriers*, Phys. Rev. Lett. **75**, 1190 (1995)

[29] G.E. Andrews, R. Askey and P. Roy, Ed. G.-C. Rota, *Special Functions (Encyclopedia of Mathematics and its Applications vol. 71)*, (Cambridge University Press, 1999)

[30] D. Chowdhury, L. Santen and A. Schadchneider, *Statistical physics of vehicular traffic and some related systems*, Physics Reports **329**, 199 (2000)

[31] D. Helbing, *Traffic and related self-driven many-particle systems*, Rev. Mod. Phys. **73**, 1067 (2001)

[32] F.L. Hall, B.L. Allen and M.A. Gunter, *Empirical analysis of freeway flow density relationships*, Transp. Res. A **20**, 197 (1986)

[33] J. Kaupužs, R. Mahnke and R.J. Harris, *Zero-range model of traffic flow*, cond-mat/0504676

[34] S.C. Benjamin, N.F. Johnson and P.M. Hui, *Cellular automata models of traffic flow along a highway containing a junction*, J. Phys. A: Math. Gen. **29**, 3119 (1996)

[35] R. Barlovic, L. Santen, A. Schadschneider and M. Schreckenberg, *Metastable states in cellular automata for traffic flow*, Eur. Phys. J. B **5**, 793 (1998)

[36] A.G. Angel, M.R. Evans and D. Mukamel, *Condensation transitions in a one-dimensional zero-range process with a single defect site*, J. Stat. Mech.: Theor. Exp. P04001 (2004)

[37] A. Rapoport and W.J. Horvath, *A study of a large sociogram*, Behav. Sci. **6**, 279 (1961)

[38] H. Jeong, B. Tombor, R. Albert, Z. Oltvai and A.-L. Barabási, *Large scale organisation of metabolic networks*, Nature **407**, 651 (2000)

[39] M.E.J. Newman, *The structure of scientific collaboration networks*, Proc. Natl. Acad. Sci. USA **98**, 404 (2001)

[40] M.E.J. Newman, *Scientific collaboration networks. I. Network construction and fundamental results*, Phys. Rev. E **64**, 016131 (2001)

[41] M.E.J. Newman, *Scientific collaboration networks. II. Shortest paths, weighted networks and centrality*, Phys. Rev. E **64**, 016132 (2001)

[42] A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert and T. Vicsek, *Evolution of the social network of scientific collaborations*, Physica A **311**, 590 (2002)

[43] A. Barrat, M. Barthélemy, R. Pastor-Satorras and A. Vespignani, *The architecture of complex weighted networks*, Proc. Natl. Acad. Sci. USA **101**, 3747 (2004)

[44] R. Albert, H. Jeong and A.-L. Barabási, *Diameter of the World-Wide Web*, Nature (London) **401**, 130 (1999)

[45] J.M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan and A.S. Tomkins, *The web as a graph: Measurements, models and methods*, in Proceedings of the International Conference on Combinatorics and Computing, Lecture Notes in Comput. Sci. **1627**, 1, Springer-Verlag, Berlin (1999)

[46] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A.S. Tomkins and E. Upfal, *The Web as a graph*, in Proceedings of the 19th symposium on Principles of Database Systems, 1 (2000)

[47] P.L. Krapivsky and S. Redner, *A statistical physics perspective on Web growth*, Computer Networks **39**, 261 (2002)

[48] B. Tadić, *Dynamics of directed graphs: the world-wide Web*, Physica A **293**, 273 (2000)

[49] A.-L. Barabási, R. Albert and H. Jeong, *Scale-free characteristics of random networks: the topology of the world-wide web*, Physica A **281**, 69 (2000)

[50] J.P.K. Doye, *Network topology of a potential energy landscape: a static scale-free network*, Phys. Rev. Lett. **88**, 238701 (2002)

[51] B. Bollobás, *Graph Theory, An Introductory Course (Graduate Texts in Mathematics 63)*, (Springer-Verlag, New York, 1979)

[52] L.C. Freeman, *A set of measures of centrality based on betweenness*, Sociometry **40**, 35 (1977)

[53] K.-I. Goh, B. Kahng and D. Kim, *Universal behaviour of load distribution in scale-free networks*, Phys. Rev. Lett. **87**, 278701 (2001)

[54] R. Pastor-Satorras, A. Vázquez and A. Vespignani, *Dynamical and correlation properties of the internet*, Phys. Rev. Lett. **87**, 258701 (2001)

[55] A. Vázquez, R. Pastor-Satorras and A. Vespignani, *Large-scale topological and dynamical properties of the Internet*, Phys. Rev. E **65**, 066130 (2002)

[56] J. Park and M.E.J. Newman, *Origin of degree correlations in the Internet and other networks*, Phys. Rev. E **68**, 026112 (2003)

[57] M.E.J. Newman, *Mixing patterns in networks*, Phys. Rev. E **67**, 026126 (2003)

[58] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins and J. Wiener, *Graph structure in the Web*, Computer Networks **33**, 309 (2000)

[59] J. Travers and S. Milgram, *An experimental study of the small world problem*, Sociometry **32**, 425 (1969)

[60] E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai and A.-L. Barabási, *Hierarchical organization of modularity in metabolic networks*, Science **297**

[61] E. Ravasz and A.-L. Barabási, *Hierarchical organization in complex networks*, Phys. Rev. E **67**, 026112 (2003)

[62] E.J. Malecki, *The economic geography of the Internet's infrastructure*, Economic Geography **78**, 399 (2002)

[63] R. Guimerà and L.A.N. Amaral, *Modeling the world-wide airport network*, Eur. Phys. J. B **38**, 381 (2004)

[64] R. Solomonoff and A. Rapoport, *Connectivity of random nets*, Bull. Math. Biophys. **13**, 107 (1951)

[65] P. Erdös and A. Rényi, *On random graphs*, Publ. Math. Debrecen **6**, 290 (1959)

[66] B. Bollobás, *Random Graphs*, (Academic Press Inc., London, 1985)

[67] A. Fronczak, P. Fronczak and J.A. Holyst, *Average path lengths in random networks*, cond-mat/0212230 (2002)

[68] M.E.J. Newman, S.H. Strogatz and D.J. Watts, *Random graphs with arbitrary degree distributions and their applications*, Phys. Rev. E **64**, 026118 (2001)

[69] S.N. Dorogovtsev, J.F.F. Mendes and A.N. Samukhin, *Metric structure of random networks*, Nucl. Phys. B **653**, 307 (2003)

[70] M. Molloy and B. Reed, *A critical point for random graphs with a given degree sequence*, Random Structures Algorithms **6**, 161 (1995)

[71] M. Molloy and B. Reed, *The size of the giant component of a random graph with a given degree sequence*, Combin. Prob. Comput. **7**, 295 (1998)

[72] F. Chung and L. Lu, *The average distances in random graphs with given expected degrees*, Proc. Natl. Acad. Sci. USA **99**, 15879 (2002)

[73] F. Chung and L. Lu, *Connected components in random graphs with given expected degree sequences*, Ann. Comb. **6**, 125 (2002)

[74] M.Á. Serrano and M.Boguñá, *Tuning clustering in random networks with arbitrary degree distributions*, cond-mat/0507535

[75] D.J. Watts and S.H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature **393**, 440 (1998)

[76] A. Barrat and M. Weigt, *On the properties of small-world networks*, Eur. Phys. J. B **13**, 547 (2000)

[77] A.-L. Barabási and R. Albert, *Emergence of Scaling in Random Networks*, Science **286**, 509 (1999)

[78] H.A. Simon, *On a class of skew distribution functions*, Biometrika **42**, 425 (1955)

[79] A.-L. Barabási, R. Albert and H. Jeong, *Mean-field theory for scale-free random networks*, Physica A **272**, 173 (1999)

[80] P.L. Krapivsky, S. Redner and F. Leyvraz, *Connectivity of Growing Random Networks*, Phys. Rev. Lett. **85**, 4629 (2000)

[81] S.N. Dorogovtsev, J.F.F. Mendes and A.N. Samukhin, *Structure of growing networks with preferential linking*, Phys. Rev. Lett. **85**, 4633 (2000)

[82] B. Bollobás, O. Riordan, J. Spencer and G. Tusnády, *The degree sequence of a scale-free random graph process*, Random Structures Algorithms **18**, 279 (2001)

[83] R. Cohen and S. Havlin, *Scale-Free Networks Are Ultrasmall*, Phys. Rev. Lett. **90**, 058701 (2003)

[84] P.L. Krapivsky and S. Redner, *Organization of growing random networks*, Phys. Rev. E **63**, 066123 (2001)

[85] S.N. Dorogovtsev and J.F.F. Mendes, *Scaling properties of scale-free evolving networks: Continuous approach*, Phys. Rev. E **63**, 056125 (2001)

[86] P.L. Krapivsky, G.J. Rodgers and S. Redner, *Degree distributions of growing networks*, Phys. Rev. Lett. **86**, 5401 (2001)

[87] R.N. Onody and P.A. de Castro, *Nonlinear Barabási-Albert network*, Physica A **336**, 491 (2004)

[88] G. Bianconi and A.-L. Barabási, *Bose-Einstein Condensation in Complex Networks*, Phys. Rev. Lett. **86**, 5632 (2001)

[89] G. Bianconi and A.-L. Barabási, *Competition and multiscaling in evolving networks*, Europhys. Lett. **54**, 436 (2001)

[90] O. Sotolongo-Costa and G.J. Rodgers, *Bose-Einstein condensation in random directed networks*, Phys. Rev. E **68**, 056118 (2003)

[91] G. Ergün and G.J. Rodgers, *Growing random networks with fitness*, Physica A **303**, 261 (2002)

[92] G. Caldarelli, A. Capocci, P. De Los Rios and M.A. Muñoz, *Scale-free networks from varying intrinsic fitness*, Phys. Rev. Lett. **89**, 258702 (2002)

[93] V.D.P. Sevedio, G. Caldarelli and P. Buttà, *Vertex intrinsic fitness: How to produce arbitrary scale-free networks*, Phys. Rev. E **70**, 056126 (2004)

[94] B. Söderberg, *General formalism for inhomogeneous random graphs*, Phys. Rev. E **66**, 066121 (2002)

[95] M. Boguñá and R. Pastor-Satorras, *Class of correlated random networks with hidden variables*, Phys. Rev. E **68**, 036112 (2003)

[96] R. Albert and A.-L. Barabási, *Topology of evolving networks: local events and universality*, Phys. Rev. Lett. **85**, 5234 (1999)

[97] S.N. Dorogovtsev and J.F.F. Mendes, *Scaling behaviour of developing and decaying networks*, Europhys. Lett. **52**, 33 (2000)

[98] G.J. Rodgers and K. Darby-Dowman, *Properties of a growing random directed network*, Eur. Phys. J. B **23**, 267 (2001)

[99] D.S. Callaway, J.E. Hopcroft, J.M. Kleinberg, M.E.J. Newman and S.H. Strogatz, *Are randomly grown graphs really random?*, Phys. Rev. E **64**, 041902 (2001)

[100] D.Y.C. Chan, B.D. Hughes, A.S. Leong and W.J. Reed, *Stochastically evolving networks*, Phys. Rev. E **68**, 066124 (2003)

[101] S.N. Dorogovtsev and J.F.F. Mendes, *Evolution of networks with aging of sites*, Phys. Rev. E **62**, 1842 (2000)

[102] S.N. Dorogovtsev and J.F.F. Mendes, *Effect of the accelerating growth of communications networks on their structure*, Phys. Rev. E **63**, 025101 (2001)

[103] N. Sarshar and V. Roychowdhury, *Scale-Free and Stable Structures in Complex Ad hoc Networks*, Phys. Rev. E **69**, 026101 (2004)

[104] R. Albert, H. Jeong and A.-L. Barabási, *Error and attack tolerance of complex networks*, Nature **406**, 378 (2000)

[105] D.S. Callaway, M.E.J. Newman, S.H. Strogatz and D.J. Watts, *Network robustness and fragility: percolation on random graphs*, Phys. Rev. Lett. **85**, 5468 (2000)

[106] R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, *Resilience of the Internet to random breakdowns*, Phys. Rev. Lett. **85**, 4626 (2000)

[107] R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, *Breakdown of the Internet under intentional attack*, Phys. Rev. Lett. **86**, 3682 (2001)

[108] P. Holme, B.J. Kim, C.N. Yoon and S.K. Han, *Attack vulnerability of complex networks*, Phys. Rev. E **65**, 056109 (2002)

[109] L. Zonghua, Y.-C. Lai and Y. Nong, *Statistical properties and attack tolerance of growing networks with algebraic preferential attachment*, Phys. Rev. E **66**, 036112 (2002)

[110] K. Klemm and V.M. Eguíluz, *Highly clustered scale-free networks*, Phys. Rev. E **65**, 036123 (2002)

[111] K. Klemm and V.M. Eguíluz, *Growing scale-free networks with small-world behaviour*, Phys. Rev. E **65**, 057102 (2002)

[112] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A.S. Tomkins and E. Upfal, *Stochastic models for the web graph*, in Proceedings of the 41st Annual IEEE Symposium on the Foundations of Computer Science, 57 , Institute of Electrical and Electronics Engineers, New York (2000)

[113] A. Vázquez, *Disordered networks generated by recursive searches*, Europhys. Lett. **54**, 430 (2001)

[114] S.N. Dorogovtsev, J.F.F. Mendes and A.N. Samukhin, *Multifractal properties of growing networks*, Europhys. Lett. **57**, 334 (2002)

[115] F. Chung, L. Lu, T.G. Dewey and D.J. Galas, *Duplication models for biological networks*, J. Comput. Biology **10**, 677 (2003)

[116] A. Vázquez, *Growing network with local rules: Preferential attachment, clustering heirarchy, and degree correlations*, Phys. Rev. E **67**, 056104 (2003)

[117] K.A. Eriksen and M. Hörnquist, *Scale-free growing networks imply linear preferential attachment*, Phys. Rev. E **65**, 017102 (2001)

[118] M.E.J. Newman, *Clustering and preferential attachment in growing networks*, Phys. Rev. E **64**, 025102 (2001)

[119] H. Jeong, Z. Néda and A.-L. Barabási, *Measuring preferential attachment in evolving networks*, Europhys. Lett. **61**, 567 (2003)

[120] E. Eisenberg and E.Y. Levanon, *Preferential attachment in the protein network evolution*, Phys. Rev. Lett. **91**, 138701 (2003)

[121] G. Palla, I. Farkas, I. Derényi, A.-L. Barabási and T. Vicsek, *Reverse engineering of linking preferences from network restructuring*, Phys. Rev. E **70**, 046115 (2004)

[122] J.D. Noh and H. Rieger, *Random walks on complex networks*, Phys. Rev. Lett. **92**, 118701 (2004)

[123] Z. Burda, J.D. Correira and A. Krzywicki, *Statistical ensemble of scale-free random graphs*, Phys. Rev. E **64** 046118 (2001)

[124] P. Bialas, Z. Burda and D. Johnston, *Condensation in the Backgammon model*, Nucl. Phys. B **493**, 505 (1997)

[125] P. Bialas, Z. Burda and D. Johnston, *Phase diagram of the mean field model of simplicial gravity*, Nucl. Phys. B **542**, 413 (1999)

[126] Z. Burda and A. Krzywicki, *Uncorrelated random networks*, Phys. Rev. E **67** 046118 (2003)

[127] J. Berg and M. Lässig, *Correlated random networks*, Phys. Rev. Lett. **89**, 228701 (2002)

[128] Z. Burda, J. Jurkiewicz and A. Krzywicki, *Network transitivity and matrix models*, Phys. Rev. E **69**, 026106 (2004)

[129] Z. Burda, J. Jurkiewicz and A. Krzywicki, *Perturbing general uncorrelated networks*, Phys. Rev. E **70**, 026106 (2004)

[130] Z. Burda, J. Jurkiewicz, A. Krzywicki, *Statistical mechanics of random graphs*, Physica A **344**, 56 (2004)

[131] G. Palla, I. Derényi, I. Farkas and T. Viscek, *Statistical mechanics of topological phase transitions in networks*, Phys. Rev. E **69**, 046117 (2004)

[132] I. Derényi, I. Farkas, G. Palla and T. Vicsek, *Topological phase transitions of random networks*, Physica A **334**, 583 (2004)

[133] I. Farkas, I. Derényi, G. Palla and T. Vicsek, *Equilibrium statistical mechanics of network structures*, in Lecture Notes in Physics **650**, 163, Springer-Verlag, Berlin (2004)

[134] C. Biely and S. Thurner, *Statistical mechanics of scale-free networks at a critical point: Complexity without irreversibility?*, cond-mat/0507670

[135] J. Park and M.E.J. Newman, *Statistical mechanics of networks*, Phys. Rev. E **70**, 066117 (2004)

[136] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, H. Teller and E. Teller, *Equation of state calculations by fast computing machines*, J. Chem. Phys. **21**, 1087 (1953)

[137] M. Baiesi and S.S. Manna, *Scale-free networks from a Hamiltonian dynamics*, Phys. Rev. E **68**, 047103 (2003)

[138] J. Berg, M. Lässig and A. Wagner, *Structure and evolution of protein interaction networks: A statistical model for link dynamics and gene duplications*, BMC Evolutionary Biology **4**, 51 (2004)

[139] *S.N. Dorogovtsev, J.F.F. Mendes and A.N. Samukhin, Principles of statistical mechanics of uncorrelated random networks*, Nucl. Phys. B **666**, 396 (2003)

[140] D. Mukamel in *Soft and fragile matter: nonequilibrium dynamics metastability and flow: proceedings of the Fifty Third Scottish Universities Summer School in Physics*, eds. M.E. Cates and M.R. Evans, (Institute of Physics Publications, 2000)

[141] O. Pulkkinen and J. Merikoski, *Phase transitions on Markovian bipartite graphs—an application of the zero-range process*, J. Stat. Phys. **119**, 881 (2005)

[142] J.D. Noh, *Stationary and dynamical properties of a zero-range process on scale-free networks*, cond-mat/0507239

[143] J.D. Noh, G.M. Shim and H. Lee, *Complete condensation in a zero-range process on scale-free networks*, Phys. Rev. Lett. **94**, 198701 (2005)

[144] S.N. Dorogovtsev, J.F.F. Mendes, A.M. Povolotsky and A.N. Samukhin, *Organization of complex networks without multiple connections*, cond-mat/0505193

[145] M. Boguñá, R. Pastor-Satorras and A. Vespignani, *Cut-offs and finite size effects in scale-free networks*, Eur. Phys. J. B **38**, 205 (2004)

[146] S. Maslov, K. Sneppen and A. Zaliznyak, *Detection of topological patterns in complex networks: correlation profile of the internet*, Physica A **333**, 529 (2004)

[147] H.J. Jensen, *Self-organized criticality: emergent complex behaviour in physical and biological systems*, (Cambridge University Press, 1998)

[148] D. Sornette, A. Johansen and I. Dornic, *Mapping self-organized criticality onto criticality*, J. Phys. I France, **5**, 325 (1995)

[149] R. Dickman, M. Muñoz, A. Vespignani and S. Zapperi, *Paths to self-organized criticality*, Braz. J. Phys. **30**, 27 (2000)

[150] G. Bianconi and M. Marsili, *Clogging and self-organized criticality in complex networks*, Phys. Rev. E **70**, 035105 (2004)

[151] A.G. Angel, M.R. Evans, E. Levine and D. Mukamel, *Critical phase in non-conserving zero-range processes and equilibrium networks*, Phys. Rev. E **72**, 046132 (2005)

[152] S.H. Yook, H. Jeong, A.-L. Barabási and Y. Tu, *Weighted evolving networks*, Phys. Rev. Lett. **86**, 5835 (2001)

[153] M.E.J. Newman, *Analysis of weighted networks*, Phys. Rev. E **70**, 056131 (2004)

[154] S.N. Dorogovtsev and J.F.F. Mendes, *Minimal models of weighted scale-free networks*, cond-mat/0408343

[155] M. Barthélemy, A. Barrat, R. Pastor-Satorras and A. Vespignani, *Characterization and modeling of weighted networks*, Physica A **346**, 34 (2005)

[156] L.R. Ford Jr. and D.R. Fulkerson, *Flows in Networks*, (Princeton University Press, 1962)

[157] M.S. Granovetter, *The strength of weak ties*, Am. J. Soc. **78**, 1360 (1973)

[158] A. De Montis, M. Barthélemy, A. Chessa and A. Vespignani, *The structure of Inter-Urban traffic: A weighted network analysis*, physics/0507106

[159] E. Almaas, B. Kovács, T. Vicsek, Z.N. Oltvai and A.-L. Barabási, *Global organisation of metabolic fluxes in the bacterium Escherichia coli*, Nature **427**, 839 (2004)

[160] J.D. Noh and H. Rieger, *Stability of shortest path lengths in complex networks with random edge weights*, Phys. Rev. E **66**, 066127 (2002)

[161] B. Derrida and H. Flyvbjerg, *Statistical properties of randomly broken objects and of multivalley structures in disordered systems*, J. Phys. A: Math. Gen. **20**, 5273 (1987)

[162] M. Barthelemy, B. Gondran and E. Guichard, *Spatial structure of the internet traffic*, Physica A **319**, 633 (2003)

[163] T. Antal and P.L. Krapivsky, *Weight-driven growing networks*, Phys. Rev. E **71**, 026103 (2005)

[164] A. Barrat, M. Barthélemy and A. Vespignani, *Weighted evolving networks: coupling topology and weight dynamics*, Phys. Rev. Lett. **92**, 228701 (2004)

[165] A. Barrat, M. Barthélemy and A. Vespignani, *Modeling the evolution of weighted networks*, Phys. Rev. E **70**, 066149 (2004)

[166] A. Barrat, M. Barthélemy and A. Vespignani, *Traffic-driven model of the World Wide Web graph*, Lect. Notes Comput. Sci. **56**, 3243 (2004)

[167] G. Bianconi, *Emergence of weight-topology correlations in complex scale-free networks*, cond-mat/0412390

[168] E. Almaas, P.L. Krapivsky and S. Redner, *Statistics of weighted treelike networks*, Phys. Rev. E **71**, 036124 (2005)

[169] M. Ángeles Serrano and M. Boguñá, *Weighted configuration model*, cond-mat/0501750

[170] K.-I. Goh, B. Kahng and D. Kim, *Nonlocal evolution of scale-free networks*, Phys. Rev. E **72**, 017103 (2005)

[171] K.-I. Goh, J.D. Noh, B. Kahng and D. Kim, *Load distribution in weighted complex networks*, Phys. Rev. E **72**, 017102 (2005)

[172] D. Zheng, S. Trimper, B. Zheng and P.M. Hui, *Weighted scale-free networks with stochastic weight assignments*, Phys. Rev. E **67**, 040102 (2003)

[173] A.G. Angel, T. Hanney and M.R. Evans, *Condensation transitions in a model for a directed network with weighted links*, cond-mat/0509238

[174] B. Hu, X.-Y. Jiang, J.-F. Ding, Y.-B. Xie and B.-H. Wang, *A weighted network model for interpersonal relationship evolution*, Physica A **353**, 576 (2005)

[175] M.R. Evans, S.N. Majumdar and R.K.P. Zia, *Factorized steady states in mass transport models*, J. Phys. A: Math. Gen. **37**, L275 (2004)